THE EFFECT OF WORKLOAD AND AGE ON COMPLIANCE WITH AND RELIANCE ON AN AUTOMATED SYSTEM

A Thesis
Presented to
The Academic Faculty

by

Sara E. McBride

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
May, 2010

THE EFFECT OF WORKLOAD AND AGE ON COMPLIANCE WITH AND RELIANCE ON AN AUTOMATED SYSTEM

Approved by:

Dr. Wendy A. Rogers, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Arthur D. Fisk
School of Psychology
Georgia Institute of Technology

Dr. Gregory M. Corso
School of Psychology
Georgia Institute of Technology

Date Approved: March 24th, 2010

ACKNOWLEDGEMENTS

I would like to thank Drs. Wendy A. Rogers and Arthur D. Fisk, the Human Factors and Aging Laboratory, and my family and friends for their support and guidance. This research was supported in part by a grant from the National Institutes of Health (National Institute on Aging) Grant P01 AG17211 under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE). This research was also supported in part by contributions from Deere & Company.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

SUMMARY

Automation provides the opportunity for many tasks to be done more effectively and with greater safety. However, these benefits are unlikely to be attained if an automated system is designed without the human user in mind. Many characteristics of the human and automation, such as trust and reliability, have been rigorously examined in the literature in an attempt to move towards a comprehensive understanding of the interaction between human and machine. However, workload has primarily been examined solely as an outcome variable, rather than as a predictor of compliance, reliance, and performance. This study was designed to gain a deeper understanding of whether workload experienced by human operators influences compliance with and reliance on an automated warehouse management system, as well to assess whether age-related differences exist in this interaction.

As workload increased, performance on the Receiving Packages task decreased among younger and older adults. Although younger adults also experienced a negative effect of workload on Dispatching Trucks performance, older adults did not demonstrate a significant effect. The compliance data showed that as workload increased, younger adults complied with the automation to a greater degree, and this was true regardless of whether the automation was correct or incorrect. Older adults did not demonstrate a reliable effect of workload on compliance behavior. Regarding reliance behavior, as workload increased, reliance on the automation increased, but this effect was only observed among older adults. Again, this was true regardless of whether the automation as correct or incorrect. The finding that individuals may be more likely to comply with

or rely on faulty automation if they are in high workload state compared to a low workload state suggests that an operator's ability to detect automation errors may be compromised in high workload situations.

Overall, younger adults outperformed older adults on the task. Additionally, older adults complied with the system more than younger adults when the system erred, which may have contributed to their poorer performance. When older adults verified the instructions given by the automation, they spent longer doing so than younger adults, suggesting that older adults may experience a greater cost of verification. Further, older adults reported higher workload and greater trust in the system than younger adults, but both age groups perceived the reliability of the system quite accurately.

Understanding how workload and age influence automation use has implications for the way in which individuals are trained to interact with complex systems, as well as the situations in which automation implementation is determined to be appropriate.

CHAPTER 1

INTRODUCTION

Technological advances have increased performance, safety, and comfort across many domains. One of the most influential technological innovations has been the rise and spread of automation. Automation is “any sensing, detection, information-processing, decision-making, or control action that could be performed by humans but is actually performed by machine” (Moray, Inagaki, & Itoh, 2000, p.44). Automation has revolutionized much of the way individuals work and go about their daily lives. Although automation originated in military and industrial applications, today’s automated systems are used in a wide array of contexts, including health care, transportation, and in the home. Just as the domains in which automation is used has expanded, the range of users that interact with automated systems has grown and expanded as well.

The importance of understanding the dynamics between humans and the automated systems they use cannot be overstated. If a designer of an automated system lacks this knowledge, this may lead users of the automation to experience decreased task performance, safety, and overall experience with the automation. All factors in the equation (i.e., the human, automation, and task environment) must be considered in the design and implementation of any automated system. It is necessary that we understand what leads users to comply with and rely on automation so that design will support optimal use.

The complexity of human-automation interaction is great. Figure 1, taken from Parasuraman and Riley (1997), depicts the numerous variables, such as trust and system reliability, that have been found or are hypothesized to influence elements of this relationship. It is clear from a brief examination of this diagram that the nature of the relationship between user, environment, and automation is complex. Research in this area has shed light on many of the variables in the model, as discussed next, but there has yet to be sufficient research examining the role of workload in the human-automation interaction. Investigating how workload may influence overall use of the automation, specifically, compliance and reliance behavior, is an important step in the understanding this intricate relationship.

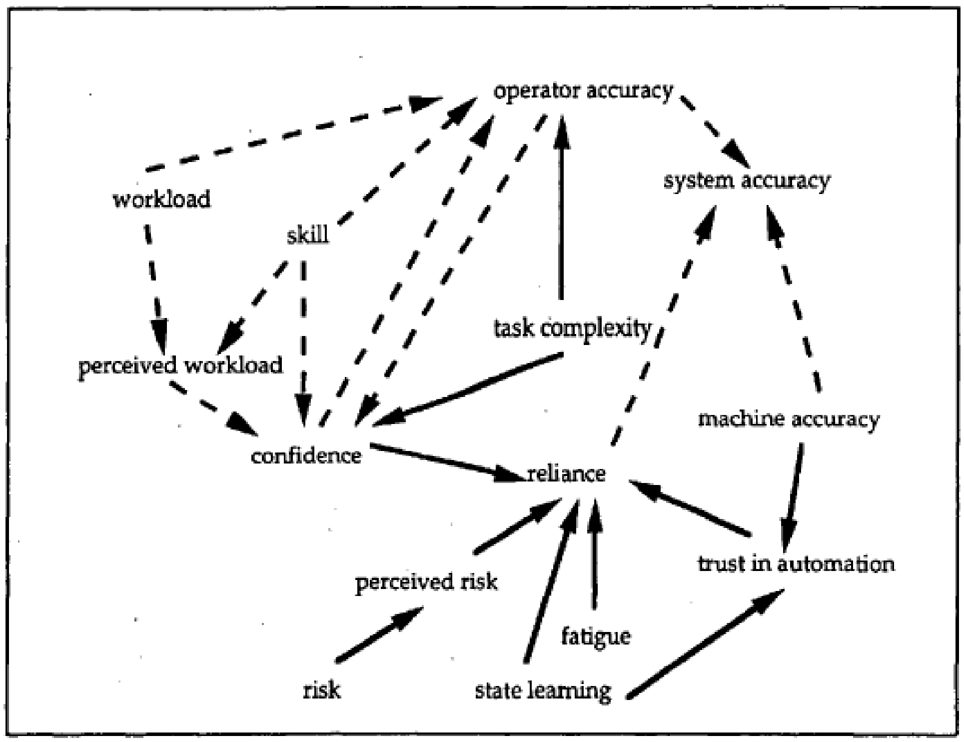


Figure 1. Dashed lines are hypothesized relationships, solid lines have empirical support (Reproduced from Parasuraman & Riley, 1997).

Use of Automation

The automation literature assesses how individuals use automation and how various factors influence the way they use automation. These factors can be thought of as relating to the automation, task, or person. Use of automation is typically examined by evaluating the constructs of compliance and reliance.

Compliance

Compliance refers to the state in which the automation presents an alert or instruction to the participant. For example, consider an in-flight collision avoidance system that can be found in the cockpit of many commercial jet liners. If the system alerts the user that a mountainside is approaching and therefore needs to pull up to avoid a collision, and the user follows the instruction and adjusts the plane's trajectory, then the user has complied. Although in this example the automation's instruction was correct and the user was correct to comply with it, it is important to consider that automation can and does make errors, and this impacts whether users should or should not comply.

If users comply with an automated system in an instance where they should not because the automation is erring, serious consequences may follow. Imagine the same automated collision avoidance system as presented in the previous example. In this case, the instruction to pull up is presented to the user too early. The user may still comply, not realizing the automation has erred, and pull up the plane at the wrong time, resulting in a collision. By examining compliance, it is possible to understand how individuals react when they are presented with a specific instruction from the automated system and how this varies as a function of whether the automation is correct or incorrect.

Reliance

Reliance refers to use of the automation during the silent state, which is any time there is not an alert from the automation present. Consider again the automated collision avoidance system. A pilot is traversing through a mountainous region with heavy fog, and relies on the collision avoidance system to provide an alert if the plane is getting too close to the surrounding mountaintop. The collision avoidance system does not alert the pilot to make any adjustments; the pilot does not do so and has therefore relied on the system.

Again, an important aspect to consider when examining reliance is whether the automation is performing as expected or making errors. If the collision avoidance system failed to alert the pilot of an approaching collision, and the pilot continued to rely on the system and not make any adjustments, the plane would likely collide with the mountain. Therefore it is important to assess whether reliance does change between situations in which the automation is correct and situations where it is incorrect.

The degree to which individuals comply with and rely on automation has been found to be influenced by a number of factors, several of which relate to the characteristics of the automation, including its reliability level, type of errors made, and the cost associated with verifying its information.

Reliability

System reliability has been found to have a significant influence on automation use. When system reliability is perfect or very high, users typically comply with it to a greater extent than if reliability is very low. Sanchez (2004) manipulated system reliability and found that as system reliability decreased, compliance decreased, and others measures such as trust and perceived reliability all dropped as well. Another study conducted by Wilkison (2008) showed that participants complied with the automation to greater extent when it was 100% accurate compared to when it was 70% accurate.

Error Type

When automation commits an error, two distinct types of errors are possible (Wickens & Carswell, 2006). A false alarm occurs when the automation incorrectly detects a signal in the environment, and a miss occurs when a signal is present in the environment, but the automation fails to detect it. Numerous studies have investigated whether the type of automation error committed has an effect on levels of compliance and reliance. Sanchez (2006) found that participants who were exposed only to false alarms tended to heavily rely on the automation, but almost never comply with the automation. Conversely, participants who only encountered misses were more likely to comply with the automation, but never relied on it.

Similarly, Johnson's (2004) work revealed that in cases where participants used an automated system that committed mostly (but not all) misses tended to rely more than participants that experienced mostly false alarms or an equal mix of false alarms and misses. This work reveals that the type of error committed by an automated system will likely lead individuals to use the automation differently.

Cost of Verification

In some scenarios, it may be possible to double-check or verify that the automated system being used is working properly and providing correct information. For instance, if one were using an automated in-vehicle navigation aid, it would be possible to verify the automation's instruction to turn left on a particular street by examining a map or a set of printed directions. However, the cost associated with this verification may vary substantially.

Ezer (2006) manipulated the cost of verification associated with an automated system by adjusting the points that were deducted for every two seconds participants spent verifying the system. She created a high cost of verification and low cost of verification and discovered that participants' compliance with the decision aid during the high cost condition was greater than during the low cost condition.

In addition to factors such as cost of verification, error type and system reliability, factors relating to the human, such as trust and workload, have been examined to understand their role in human-automation interaction.

Trust

Trust in automation has been identified as a predictor of compliance and reliance behavior in numerous studies, especially in instances where the task is complex or where a thorough understanding of the system is not possible (Lee & See, 2004; Muir, 1994; Parasuraman & Riley, 1997). Trust has been described in many ways, but the elements that appear central to the notion of trust include the trustor's willingness to place him/herself in a vulnerable position and an expectation that the trustee will act in a particular, beneficial manner (Lee & See, 2005).

Operator Workload

Workload has been thought of as the supply and demand of attentional or processing resources (Tsang & Vidulich, 2006). Two main determinants of workload have been identified; the exogenous task demands and the endogenous supply of attentional or processing resources to support activities such as information processing, memory, and decision making. The exogenous task demands includes factors such as task difficulty and task priority, whereas the endogenous supply of resources might be moderated by individual differences (Tsang & Vidulich).

Assessing an individual's level of workload has been accomplished by evaluating the individual's subjective experience of workload, performance, or physiological manifestations of workload. Subjective measures typically are multidimensional in nature to support the diagnosticity of the tool so that participants can be more specific in describing the various aspects of their experience (Tsang & Vidulich). Performance measures of workload usually assess performance on the primary as well as secondary task, if a secondary task exists. Physiological measures of workload include cardiovascular (heart rate and heart rate variability), ocular (pupil dilation and eye movements), and brain activity measures (EEG, ERP, PET, fMRI).

In the automation literature, workload has often been examined as a dependent measure. That is, one of the great advantages of implementing automation in complex systems was that it was believed to reduce operator workload. That is, users would be able to devote more of their attention to certain tasks because other tasks that had previously demanded attention would now be automated. This would lead to a decrease in operator errors and improvement in overall performance. Although this aspect of human-automation interaction is certainly worth examining, workload has not received enough attention for its role in determining how individuals use automation and how this use consequently affects task performance. The literature that does exist that has examined the role of workload as a predictor of compliance, reliance, and performance have demonstrated mixed results, which may be the result of several issues. In many studies, workload has not been systematically measured and compliance and reliance were not assessed, just performance.

One study examined performance in unmanned aerial vehicle (UAV) simulated missions in which perfectly reliable automation supported pilots in either a low or high workload situation (Dixon & Wickens, 2005). These varying levels of workload were attained by manipulating the number of UAVs each pilot was responsible for monitoring (i.e., manipulating the task demands); a single UAV in the low workload condition and two UAVs in the high workload condition. Their data suggested that detection time for a system failure was significantly longer for participants experiencing high workload compared to low workload. Additionally, memory failures were more common among participants in the high workload situations.

However, the automation implemented in this experiment was perfectly reliable. Automation is not always 100% reliable. Examining these factors in situations where a human must interact with less than perfect automation may be more useful, as unreliable automation could be less beneficial or even detrimental to performance in situations where workload is high and operators are only further burdened by having to monitor the automation.

An earlier study sought to examine how observer compliance with an imperfect automated system correlated with task difficulty (Maltz & Shinar, 2003). They found that when task difficulty was high, compliance with the automation increased, resulting in performance superior to participants who did not have an automated aid. However, when task difficulty was low, high levels of reliance on the automation actually impaired performance. That is, when task difficulty was low, participants performed better if they did not use the automated aid. This study highlights the importance of understanding the burden of the task on the user or operator before making the decision to implement automation, as in some cases, performance may actually be hindered by introducing automation.

Biros, Daly, and Gunsch (2004) examined the role of task workload and trust in the use of an automated system. They discovered that when participants had low automation trust, individuals tended to comply with automation more when task load increased. The authors postulated that as the task load increases and more environmental cues are present in the environment, a decision-maker may resort to using the automation as a means of keeping up with the environment, and that this may be especially concerning in cases where inaccurate information is being presented.

However, in contrast to the findings just discussed, Riley (1994, as cited in Parasuraman & Riley, 1997) found no significant difference in automation usage after manipulating the difficulty of a tracking task in a dual-task situation. A possible explanation of these findings may be that the young adults being tested simply preferred manual control over automated control. To investigate this possible confound, the study was replicated with the participants consisting of pilots, and there was again no relationship between task difficulty and automation usage. However, it is possible that Riley's manipulation of difficulty was not strong enough to actually increase or decrease workload, as performance did not vary between the different groups.

One very important point to consider with the studies highlighted here is that workload experienced by the participant was not always assessed. By manipulating characteristics of the task and environment, the researchers may inappropriately assume that workload experienced by the user is changing in the way they predict. It is a mistake to draw conclusions about how varying levels of workload affect compliance, reliance, and performance without ascertaining workload with valid measures, especially when other factors are also being manipulated. Thus, in the present study subjective workload and performance were assessed to draw conclusions about workload experienced by participants.

Age-related Differences

The definition for workload provided earlier described two aspects of workload; one related to exogenous factors such as task demands, and another related to the endogenous supply of attentional and processing resources. Although the former component of workload has received some attention, as evidenced by the literature review just presented, not enough research has been conducted to understand the latter component. Specifically, research that compares individuals with varying levels of attentional and processing resources is needed. Two groups where this difference would likely be present are younger and older adults.

Age- related differences have been well documented for cognitive processes such as working memory and attention, which might indicate that older adults would experience higher workload that younger adults (e.g., Braver & West, 2008; Kramer & Madden, 2008). If workload leads to differences in compliance, reliance, and performance, then these differences might also appear when comparing younger and older adults in the same task. Further, automated systems are often implemented in situations comprising of dual or multiple tasks, which might be particularly beneficial to older adults, who tend to have more difficulties performing two tasks concurrently compared to younger adults (e.g., Kramer & Madden, 2008). However, of the minimal studies investigating whether age-related differences exist in compliance and reliance behavior, the findings have been mixed. A clear understanding of older adults' interactions with automation will be necessary to design and implement automation that they will be able to benefit from.

Studies have found that, overall, older adults typically comply with automation to a greater extent than younger adults (e.g., Johnson, 2004; Mayer, 2008). The explanation given for this finding is that older adults are experiencing greater workload and therefore must use the automation to a greater extent. However, it is not clear as to whether increased workload necessarily leads to increased compliance behavior.

In fact, some research has shown that when older adults report higher subjective workload than younger adults, their compliance behavior is no different from that of younger adults. For example, Ho, Wheatley, and Scialfa (2005) asked younger and older adults to perform a dual task in which they would be responsible for solving simple math problems as well as taking prescribed amounts of various medications at the appropriate times using a computer simulation of the automated medication management system. The participants were assessed in terms of their trust of the automated decision aid, self-confidence, general computer anxiety, and subjective workload. The results indicated that older adults, compared to younger adults, reported greater trust in the automated medication management system. Older adults also rated their workload higher than younger adults. However, older adults did not comply with the automation differently than the younger adults.

Understanding how workload influences older adults' use of automation is as unclear as research conducted with younger adults. Workload is often the explanation researchers give when older adults are found to comply with automation to a greater degree. However, this account does not have a basis in experimental findings. Workload must be systematically manipulated to determine whether high workload does indeed lead individuals to use automation to a greater degree. Considering the numerous areas in which automation has the potential to provide great benefits to the older adult, such as in-home medical devices, a comprehensive understanding of how older adults approach automation should be a priority.

Study Overview

The purpose of this research was to ascertain what effect workload has on compliance with and reliance on an automated aid, and whether younger and older adults complied with and relied on the automation differently. This was accomplished by systematically manipulating and measuring workload. The current study employed a simulated automated warehouse management system (AWMS). The AWMS operated in the context of dual-task scenario in which participants played the role of a warehouse manager who was responsible for correctly receiving packages coming into the warehouse, and for dispatching full trucks out of the warehouse.

This simulation was chosen because it is representative of a class of automated systems that have several characteristics. First, these systems are dynamic. They are not static environments but involve changing elements that must be continuously monitored and assessed. The dual-task nature of this simulation affords the ability to determine how changing certain components of one task in the system affects use of the automated aid.

Additionally, the novelty of the tasks involved in the simulation was an important feature, as it ensured that differences found between participants or groups could not be attributable to prior experience with similar tasks.

CHAPTER 2

METHOD

Participants

Participants consisted of 42 younger adults between the age of 18 and 28 and 42 older adults between the age 65 and 75 participated in this study. The younger adults were Georgia Institute of Technology undergraduates who received three hours of Experimetrix credit. The older adults were recruited from the Human Factors and Aging Laboratory database and received twelve dollars per experimental hour.

Ability Tests

Participants' near and far vision were assessed using the Snellen visual acuity exam. All participants demonstrated at least 20/40 near and far vision. Additionally, the Reverse Digit Span, Digit Symbol Substitution (Wechsler, 1997), and Shipley Vocabulary tests (Shipley, 1986) were administered as measures of memory span, perceptual speed, and verbal ability, respectively. These abilities tests were administered to assess whether groups differed between the workload conditions, as well as whether age-related differences existed.

Materials

Simulated Automated System

The participants interacted with an Automated Warehouse Management System (AWMS) that required them to act as a warehouse manager in charge of two tasks: 1) receiving packages into inventory; and 2) dispatching trucks once they were filled to capacity.

In the receiving packages task, participants were presented with a barcode consisting of several symbols. Beside this individual barcode, the participants were presented with a list of barcodes (see Figure 2 for a screen shot of the Receiving Packages task). Their goal was to find the matching barcode in the list. They navigated through the list of barcodes using the up and down arrow keys and selected a barcode using a key labeled "Receive". Younger adults had only seven seconds to make a selection; older adults received ten seconds. Older adults received one and a half times the amount of time to compensate for decreases in attention-switching ability and general cognitive slowing (Craik & Salthouse, 2000).

Feedback for this task appeared in three forms. If the correct barcode was selected, feedback indicating a correct response and the points earned was presented. If the wrong barcode was selected, they received feedback indicating an incorrect response and the points lost. If they did not select a barcode before the time ran out, feedback indicating that time ran out and the points lost were presented. After the feedback was presented in each of these three cases, a new barcode and a new list of barcodes were automatically presented, and the participant continued the task.

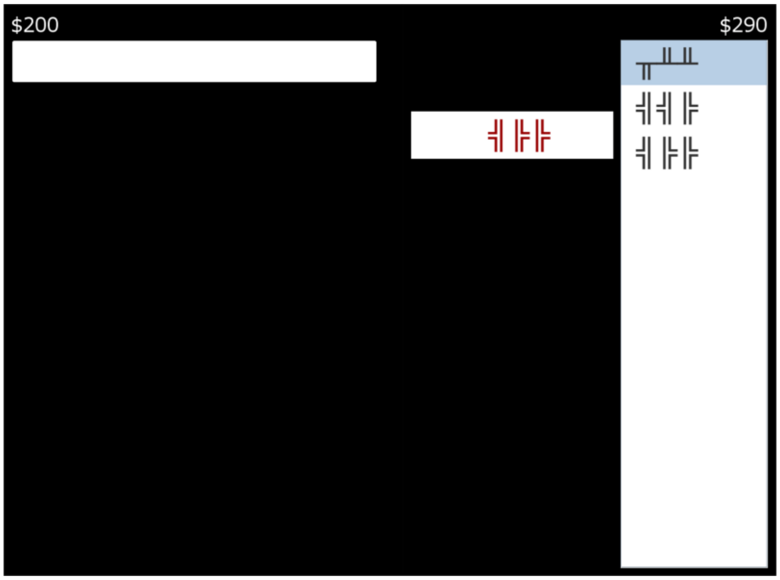


Figure 2. Screen shot of the Receiving Packages task.

While participants were completing the receiving packages task, they were also responsible for dispatching full trucks on time using the help of an automated aid that monitored the interior of the truck (see Figure 3 for a screen shot of the Dispatching Trucks task). When the automation detected that the truck had reached full capacity, it sent a message to participants alerting them to dispatch the truck. Participants could then decide to dispatch the truck by pressing a key labeled “Dispatch Truck” or to verify the automation’s suggestion and view the interior of the truck.

To view the interior of the truck, participants were required to press and hold down the spacebar. This action was given at a cost. When the spacebar was pressed, the receiving packages task disappeared from the screen, keeping the participants from being able to perform that task. All keys for the receiving packages task as well as the “Dispatch Truck” key were disabled while the truck interior was being viewed. Additionally, when the spacebar was initially pressed, there was a two second delay before the interior of the truck would appear, increasing the amount of time being taken away from the receiving packages task. Participants were able to view the interior of the truck at any time, not only when the automation gave an alert.

After the truck reached its full point, it continued to be filled with boxes until the participant dispatched it. If the participant did not dispatch it within ten seconds of it becoming full, the truck would become overloaded, meaning that it was filled beyond capacity. If participant failed to dispatch the truck before it overloaded, or if they dispatched the truck before it was full, points were deducted from the dispatching trucks tasks. The time until a truck reached its full point varied between 12 and 22 seconds.

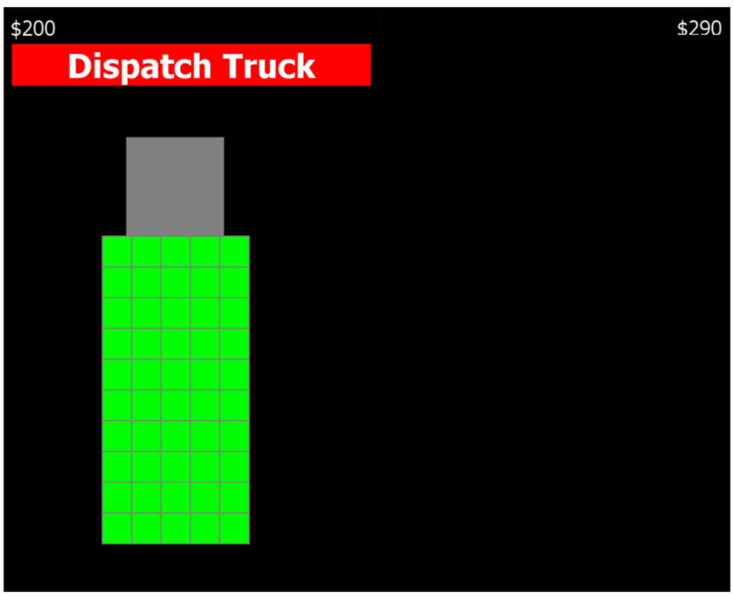


Figure 3. Screen shot of the Dispatching Trucks task. More specifically, this represents what was seen when viewing the truck.

The simulated automated system was capable of making errors. It could commit two types of errors, a miss or a false alarm. A miss constituted the automation not recognizing that the truck was full and, consequently, failed to send an alert. A false alarm occurred when the system notified the participant that the truck was full before it actually was.

General Trust in Automation Questionnaire

Prior to any exposure to the automated system, participants completed a questionnaire designed to assess trust in a hypothetical automated system. This served as a baseline trust measure. This measure was developed from questionnaires used by Mayer (2008), Sanchez (2006), Johnson (2004), and Jian, Bisantz, and Drury (2000) and is presented in Appendix A.

Interim Questionnaire

After the completion of each block of trials, participants were asked to rate the reliability of the automation, their trust in the automation, perceived compliance with the automation, perceived reliance on the automation, and subjective workload (see Appendix B). Subjective workload was measured using modified versions of the NASA-TLX workload assessment and the Subjective Workload Assessment Technique (Hart & Staveland, 1988; Reid & Nygren, 1988). Only the six individual scale ratings were administered from the NASA-TLX, and only the three subjective scales were used from the SWAT.

Two measures of subjective workload were administered to provide multiple lines of evidence that workload was being effectively manipulated between groups. The questionnaire, excluding the subjective workload questions, was also administered at the start of the second day of testing. The participants were asked about their overall experience with the automated system over the course of the entire first day of testing, to assess whether any shifts in their opinions or attitudes occurred during the period of time from the end of Day 1 to the start of Day 2.

Strategy and Performance Questionnaire

Upon completion of all experimental blocks, participants completed a strategy and performance questionnaire (see Appendix C). Questions assessed participants' attitudes regarding the difficulty of the task, strategies used to accomplish the tasks, and whether these changed during the course of the experiment.

Procedure

Figure 4 presents the experimental procedure. Participants were first presented with an informed consent document. Next they completed the Snellen near and far visual acuity exam, the Shipley Vocabulary test, and the General Trust in Automation questionnaire. Participants were then given specific information and instructions about the automated warehouse management system. Next, seven practice blocks were completed to allow the participants to familiarize themselves with the receiving and dispatching tasks.

The first and second practice blocks consisted of only the Receiving Packages task, first without the time limit of seven or ten seconds enforced, and then with the appropriate time limit enforced in the next block. In each of these first two practice blocks, participants were required to reach a performance criterion of five times the points awarded for a correct response before they were permitted to move onto the next practice block.

The third block consisted of only the dispatching task and participants were exposed to five trucks, but did not have to reach any performance criterion. The fourth block contained three trucks, all of which included a false alarm by the system. The fifth block consisted of another three trucks to the participant, but each of those involved the system committing a miss. The sixth block involved both the receiving and dispatching tasks, and the automation performed without error. On this block, the participant was presented with five full trucks and did not have to reach a performance criterion to move on to the experimental blocks. The final block consisted of five trucks, two of which included an error by the automation.

When the practice trials were finished, the participant began data collection. Participants were responsible for dispatching 20 trucks per block, and they completed a total of eight blocks. Due to concerns regarding fatigue, the eight blocks were completed over the course of two sequential days rather than in one single day, with four blocks occurring on each day. After each block, the participant completed the interim questionnaire. The AWMS's reliability was consistent at 70% across blocks. Within each block, three misses and three false alarms occurred.

On the second day of testing, participants completed the Reverse Digit Span and Digit Symbol Substitution tests. They were then given the same information and instructions regarding the automated system, but only completed a single practice block that consisted of both the Receiving Packages and Dispatching Trucks tasks. Upon completion of these blocks, participants completed the strategy and performance questionnaire, were debriefed, compensated, and thanked for their participation.

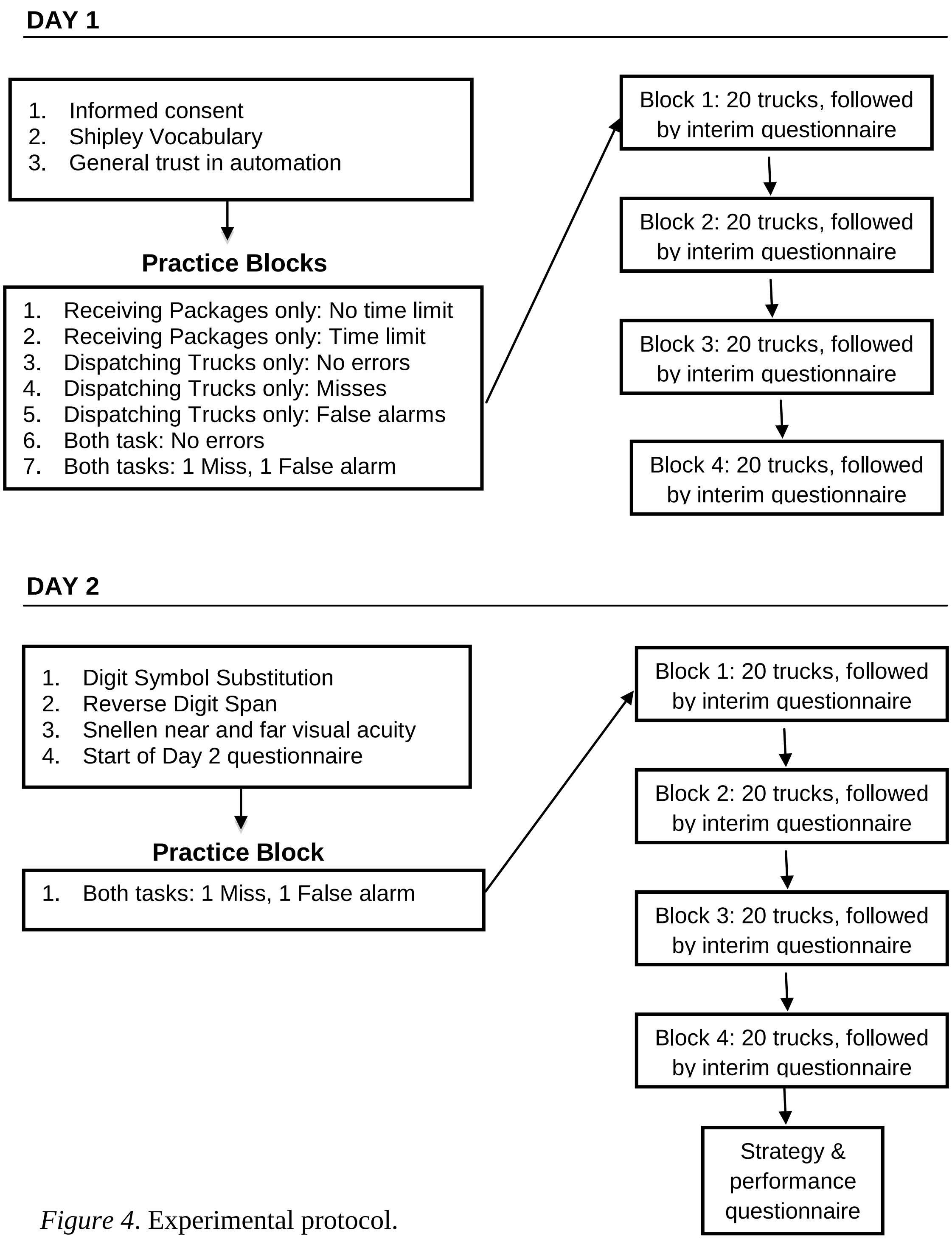


Figure 4. Experimental protocol.

Design

Age served as a grouping variable (young or old) and level of workload (low, moderate, and high) served as a between-participant variable. Participants were randomly assigned to a workload condition (high, moderate, or low).

Independent Variables

To manipulate the level of workload experienced by the participants, the number of characters in each barcode and the number of barcodes in the list of possible matches varied in the Receiving Packages task. The levels of workload presented were selected because they evoked the desired level of workload during pilot testing as evidenced by scores on the NASA-TLX as well as performance on the task (see Table 1 and Appendix D).

Table 1

Workload Manipulation for Younger and Older Adults

Younger adults				Older adults			
	Low	Moderate	High		Low	Moderate	High
Characters per barcode	3	4	6	Characters per barcode	1	3	4
Barcodes in list	3	6	11	Barcodes in list	2	3	6

Dependent Variables

Reliance and compliance were calculated by assessing the percentage of trucks that were not viewed during the non-alarm state and during the alarm state, respectively. Performance was measured on both tasks in terms of proportion correct and incorrect. Subjective measures were included from the interim questionnaire, including perceived reliability, trust, and subjective workload.

Point Structure

Participants began each experimental block with zero points for both the Receiving Packages and Dispatching Trucks task. Because it was important for participants to consider both tasks equally, the points attainable on the two tasks were designed to be roughly equivalent. If participants had been able to receive substantially more points by focusing on only one task, this is likely what they would have done. However, because the workload varied on the Receiving Packages task, making the task easier or more difficult, it was necessary to reduce the points that could be attained when the task was easier, and increase the points attainable when the task was more difficult. This allowed for the points attainable to be consistent across the workload groups. Therefore, the point scheme for younger and older adults is shown in Table 2.

Table 2

Receiving Packages Task Point Structure

	Low	Moderate	High
Older adults	15	20	25
Younger adults	10	15	20

In the Dispatching Trucks task, participants were awarded 100 points if they dispatched a full truck. Participants were deducted 200 points if they dispatched an under-filled truck, or if they failed to dispatch the truck in time, causing it to overload. The variation in points allotted/deducted across the two tasks existed because pilot testing conducted by Mayer (2008) revealed that if points were of a similar magnitude in the two tasks, participants paid more attention and gave more effort to the Receiving Packages task in which points were earned at a much faster pace than in the Dispatching Trucks task. Several participants noted that the points they lost from not correctly dispatching a truck could be made up by focusing on the Receiving Packages task, and that it just felt "easier" to approach the dual task this way, even though they were repeatedly informed that both tasks were equally important. Because this strategy undermines the participant's compliance with and reliance on the automated system, the point structure for the tasks was devised so that focusing on a single task was discouraged.

CHAPTER 3

RESULTS AND DISCUSSION

Overview of Analyses

Statistical tests were conducted using a repeated measures analysis of variance (ANOVA) and planned contrasts. The alpha level was set to .05 for all statistical tests unless otherwise noted. Error bars in all figures represent standard error of the mean. Data were averaged across blocks because patterns were similar across blocks. Block level data are presented in Appendix E. Data collected from the interim questionnaire administered at the start of the second session and strategy and performance questionnaire are not presented.

Ability Tests and Demographics

The Reverse Digit Span, Digit Symbol Substitution (Wechsler, 1997) and Shipley Vocabulary (Shipley, 1986) tests were administered to determine whether groups differed across the workload conditions within each age group, as well as whether there were differences between younger and older adults.

Table 3 depicts the means and standard deviations for the ability tests as well as demographic data for each workload group within each age group. No significant differences were found between workload groups for younger or older adults, all *p*s > .15. Younger adults outperformed older adults on both the Digit Symbol Substitution and Reverse Digit Span test, $F(1, 78) = 92.49$, $p < .01$, $\eta^2 = .53$, $F(1, 78) = 6.19$, $p < .01$, $\eta^2 = .07$, respectively. Older adults had significantly higher scores than younger adults on the Shipley Vocabulary test, $F(1, 78) = 10.28$, $p < .01$, $\eta^2 = .11$. All participants scored within three standard deviations for their group on the ability tests; therefore, no participants were excluded.

Table 3

Means (Standard Deviations) for the Demographic and Ability Data

	Low workload	Moderate workload	High workload	Overall
Younger adults				
Age	19.69 (1.44)	19.86 (2.11)	21.57 (2.98)	20.39 (2.39)
Education (years)	13.43 (1.22)	13.14 (1.03)	13.86 (1.23)	13.48 (1.17)
Health[a]	4.14 (0.66)	4.07 (0.62)	4.00 (0.55)	4.07 (0.60)
Health compared to others[a]	3.93 (0.83)	4.14 (0.77)	3.71 (0.61)	3.93 (0.75)
Digit Symbol Substitution[b]	78.00 (13.91)	70.64 (8.05)	74.36 (10.85)	74.33 (11.33)
Reverse Digit Span[b]	11.29 (2.13)	10.57 (1.79)	9.86 (2.32)	10.57 (2.12)
Shipley Vocabulary[c]	30.71 (4.95)	32.14 (2.88)	30.93 (2.70)	31.26 (3.62)
Older adults				
Age	70.71 (2.70)	69.86 (2.80)	69.93 (3.02)	70.17 (2.80)
Education (years)	14.07 (2.43)	15.07 (2.09)	17.36 (2.37)	15.50 (2.64)
Health[a]	3.54 (0.88)	3.92 (0.76)	3.71 (0.73)	3.73 (0.78)
Health compared to others[a]	3.92 (0.95)	4.46 (0.66)	4.00 (0.78)	4.13 (0.82)
Digit Symbol Substitution[b]	51.64 (13.61)	47.57 (9.96)	52.71 (10.20)	50.64 (11.32)
Reverse Digit Span[b]	8.14 (3.74)	10.21 (2.67)	9.21 (2.15)	9.19 (2.98)
Shipley Vocabulary[c]	33.36 (6.08)	33.00 (6.34)	36.93 (2.37)	34.43 (5.43)

Note:
[a] 1 = poor, 2 = fair, 3 = good, 4 = very good, 5 = excellent
[b] Wechsler (1997) – score is number correct
[c] Shipley (1986) – score is number correct

Receiving Packages

In the Receiving Packages task, participants were required to find a match in a list for a given barcode within a restricted amount of time. Participants either matched the barcode correctly, incorrectly, or timed out, meaning they failed to make any response within the allotted time. Depending on what workload group the participant belonged to, the list of possible matches presented to them varied in length and complexity. Performance on this task was examined to understand how this workload manipulation affected participants.

Receiving Packages Performance

As the workload of the Receiving Packages task increased, performance on the task declined for both younger and older adults (see Figure 5 & Figure 6). A main effect of workload group was found for the percentage of Receiving Packages trials correctly matched, revealing that groups experiencing a higher workload correctly matched fewer barcodes than groups with lower workload, $F(2, 78) = 41.08$, $p < .01$, $\eta^2 = .06$.

The two types of errors participants could make in this task (incorrect match or timeout) were also examined to determine whether the workload manipulation affected one type of error or both types of errors. The analysis showed that both types of errors were influenced by the workload manipulation, in that the percentage of incorrect trials and timeout trials both increased with increasing workload, $F(2, 78) = 12.93$, $p < .01$, $\eta^2 = .24$, $F(2, 78) = 37.43$, $p < .01$, $\eta^2 = .46$ (see Figure 5 & Figure 6).

For the younger adult group, depicted in Figure 5, planned contrasts between the workload groups revealed significant differences for the percentage of trials that were correct, incorrect, and timeout. The low and moderate workload groups had a significantly higher percentage correct than younger adults in the high workload group, $t(39) = 10.11$, $p < .01$, $t(39) = 28.28$, $p < .01$, respectively. There was not a significant difference between younger adults in the low and moderate workload groups, $p = .08$.

For incorrect matches, planned contrasts revealed a significant difference between the low and high, $t(39) = -5.17$, $p <.01$, and moderate and high workload groups, $t(39) = -4.05$, $p <.01$, but not between the low and moderate groups, $p > .27$. As for trials in which the younger adults timed out, a similar pattern was observed in which significant differences were found between the low and high, $t(39) = -10.55$, $p <.01$, and moderate and high workload groups, $t(39) = -8.86$, $p <.01$, but not between the low and moderate groups, $p = .10$.

Older adults' performance on the Receiving Packages task is presented in Figure 6. As previously mentioned, a main effect of workload was found for the percentage of trials that were correct, incorrect, and timeout. As the workload increased, the percentage of correct trials decreased, and the percentage of error trials (both incorrect and timeout) increased. For the correct trials, significant differences were found between all three workload groups, such that the low and moderate workload groups had a higher percentage correct than older adults in the high workload group, $t(39) = 4.55$, $p <.01$, $t(39) = 2.55$, $p = .02$, and older adults in the low groups had a higher percentage correct than the moderate workload group, $t(39) = 2.00$, $p = .05$.

As for incorrectly matched trials, older adults in the high group incorrectly matched significantly more than the low group, $t(39) = -2.495$, $p = .02$. No other contrasts were significant, all *p*s > .07. Examining the timeout trials, significant differences were found between all workload groups, the low and high, low and moderate, and moderate and high, $t(39) = -4.36$, $p < .01$, $t(39) = -2.35$, $p = .02$, $t(39) = -2.02$, $p = .05$.

These data demonstrate that the workload manipulation significantly increased the percentage of errors committed in the Receiving Packages task, including incorrect matches and as well as timeouts, thereby driving down the percentage of correct trials for both younger and older adults.

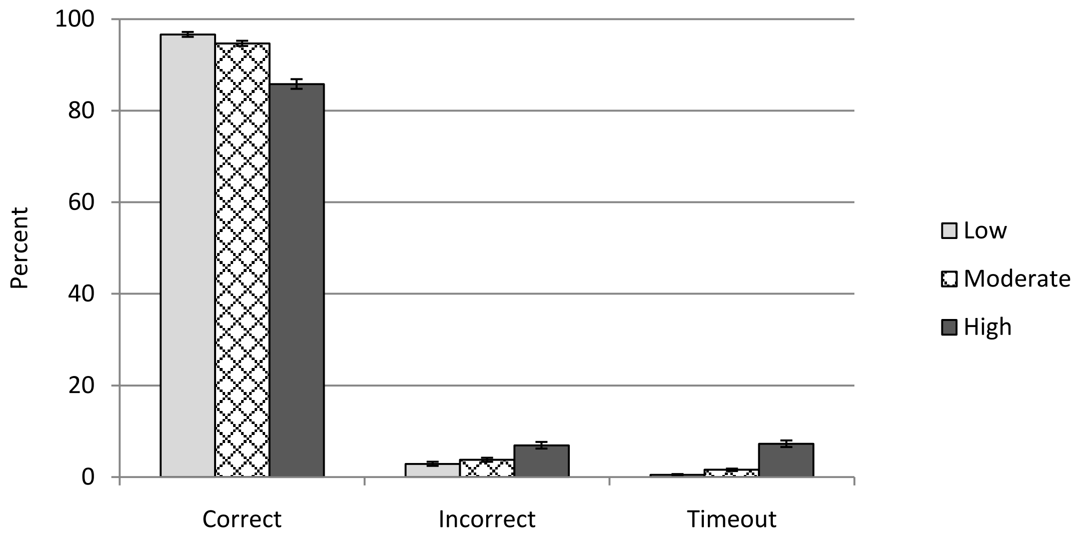


Figure 5. Younger adults' performance on the receiving packages task, divided by workload group.

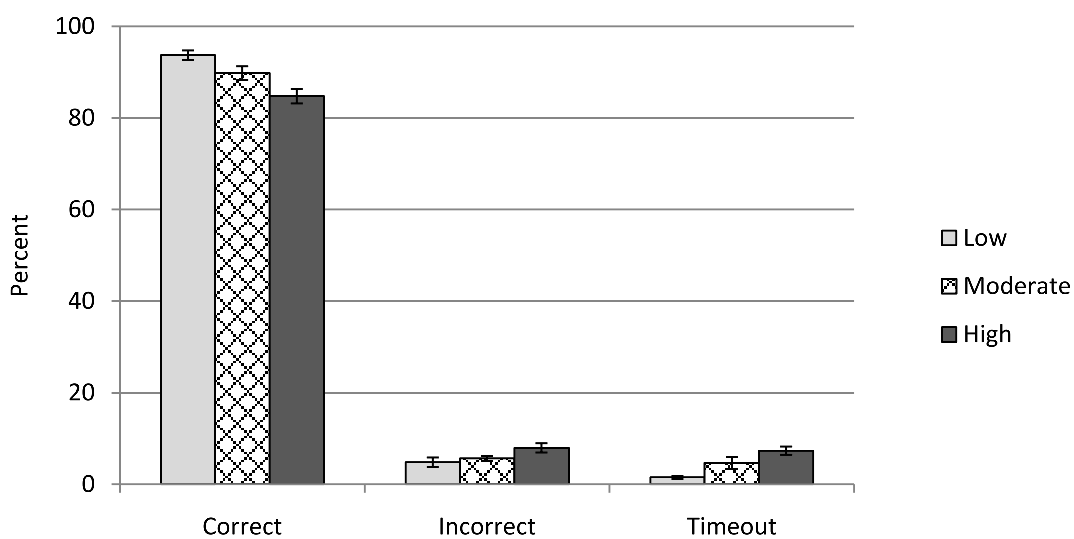


Figure 6. Older adults' performance on the receiving packages task, divided by workload group.

In addition to effect of workload on performance in the Receiving Packages task, differences between the younger and older adults were also found, with younger adults outperforming older adults on this task (see Figure 7). A main effect of age was found for the percentage of trials that were correct, demonstrating that younger adults correctly matched significantly more barcodes than older adults, $F(1, 78) = 10.58$, $p < .01$, $\eta^2 = .06$. To understand the types of errors participants were making in the task, both incorrect and timeout trials were examined to determine whether the age groups differed in both or just one type of error. Older adults made significantly more incorrect matches and also had more timeout trials than younger adults, $F(1, 78) = 6.95$, $p = .01$, $\eta^2 = .06$, $F(1, 78) = 5.25$, $p = .03$, $\eta^2 = .03$, respectively. The workload by age interaction was not statistically significant for percentage correct, incorrect, or timeout, all *p*s > .13.

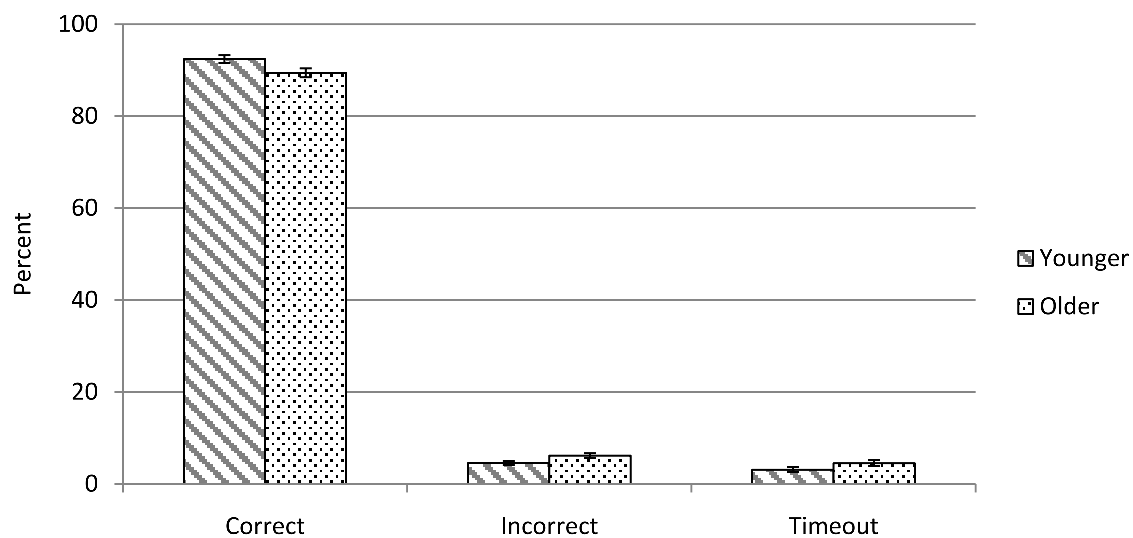


Figure 7. Percent of receiving packages trials that were correctly matched, incorrectly matched, or timed out.

Dispatching Trucks

One goal of this study was to understand how participants interacted with the automated aid that was present in the Dispatching Trucks task while simultaneously performing the Receiving Packages task under different levels of workload. To investigate this, the Dispatching Trucks task was examined in terms of performance, as well as the degree to which the participant complied with and relied on the automated system. Further, detailed information regarding the instances where the participant did not comply or rely was analyzed. Finally, the participants' subjective experiences using the automation were considered by examining their ratings of trust in the automation as well as their perceptions of reliability of the system.

Dispatching Trucks Performance

In the Dispatching Trucks task, participants either dispatched a truck correctly, meaning the truck was dispatched on time when it was appropriately full, dispatched the truck when it was not full, or failed to dispatch the truck in time resulting in an overloaded truck. These data were analyzed to determine whether workload had an effect on performance in the Dispatching Trucks task. There was not a main effect of workload for the percentage of trucks dispatched correctly or overloaded, all *p*s > .12, but there was a significant main effect of workload on the percentage of trucks that were dispatched when they were not full, $F(2, 78) = 3.77$, $p = .03$, $\eta^2 = .08$. Although a consistent pattern was not observed across both age groups, as was the case with the Receiving Packages data, the younger adult data revealed that as workload increased, performance on the Dispatching Trucks task decreased.

Younger adults' performance in the Dispatching Trucks task is depicted in Figure 8. The data show that as workload increased, the percentage of trucks correctly dispatching decreased. Specifically, younger adults in the high workload group had a significantly lower percentage correct than the low workload group, $t(39) = 2.63$, $p = .01$. There was not a significant difference between younger adults in the low and moderate workload groups or moderate and high workload groups for percentage correct, all $p > .07$.

Workload also impacted the likelihood of a younger adult dispatching a not full truck. There was a main effect of workload, and planned contrasts showed younger adults in the high workload group dispatched significantly more not full trucks than young adults in the low group, $t(39) = -2.65$, $p = .01$. No significant differences were found between the low and moderate or moderate and high groups, all *p*s > .07. Finally, although the pattern seen in the overloaded trucks data is consistent with the not full trucks data, there were no statistically significant differences between any of the workload groups, all *p*s > .06.

Older adults' performance in Dispatching Trucks is presented in Figure 9. Planned comparisons did not reveal any statically significant effects of workload on correct, not full, or overloaded trucks, all *p*s > .18.

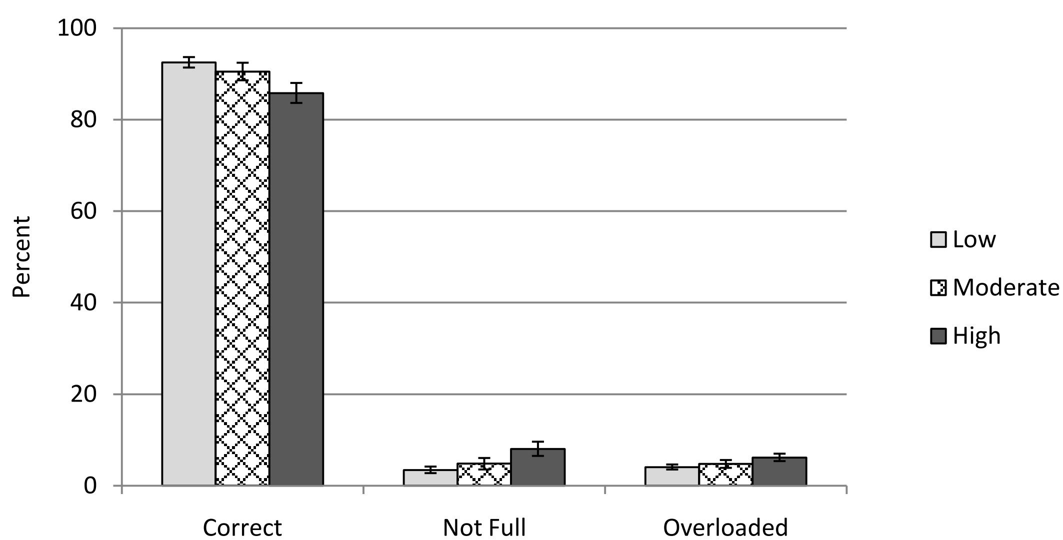


Figure 8. Younger adults' performance on the dispatching trucks task, divided by workload group.

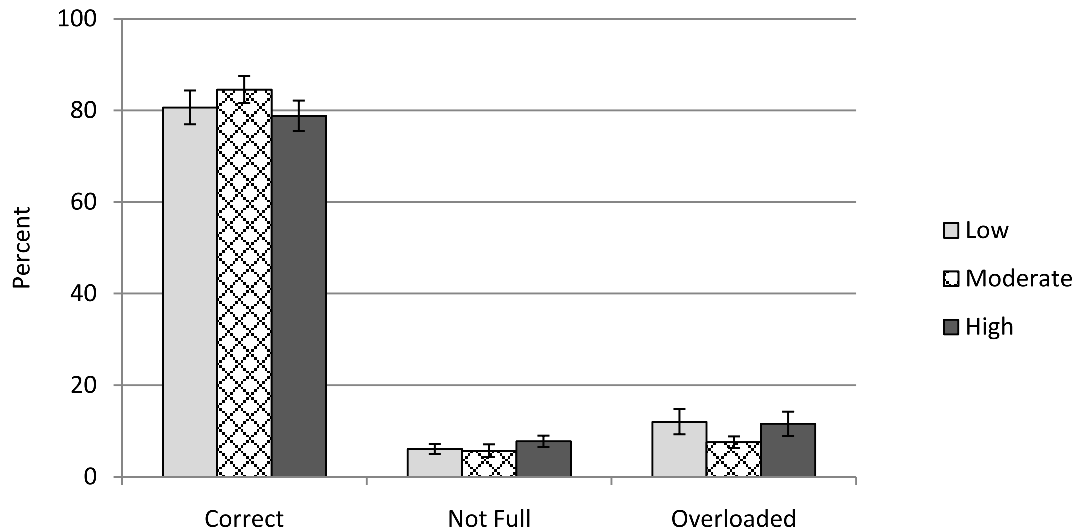


Figure 9. Older adults' performance on the dispatching trucks task, divided by workload group.

In addition to understanding how the workload of the Receiving Packages task affected performance in the Dispatching Trucks task, another aim of analyzing participants' performance in the Dispatching Trucks task was to determine whether age-related differences existed. Figure 10 depicts younger and older adults' performance on Dispatching Truck, and the analysis revealed that younger adults had a significantly higher percentage of trucks that were dispatched correctly, and they also had significantly fewer trucks that overloaded, $F(1, 78) = 14.29$, $p < .01$, $\eta^2 = .15$, $F(1, 78) = 14.58$, $p < .01$, $\eta^2 = .15$, respectively. However, younger and older adults did not differ significantly in the percentage of trucks that were dispatched when not full, $p > .29$. The age by workload group interaction was not significant for percentage of trucks that were correct, not full or overloaded, all *p*s > .34.

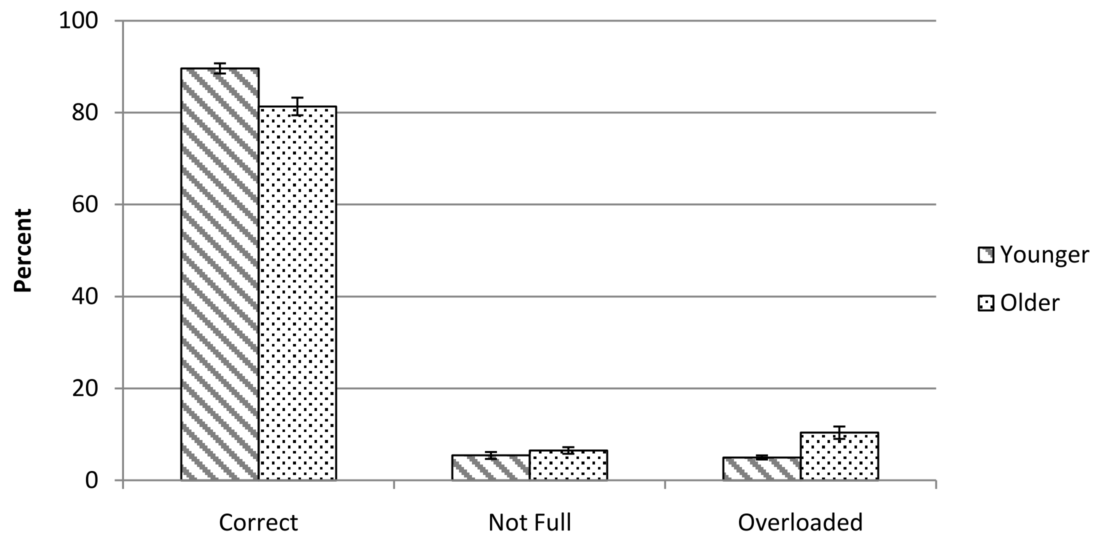


Figure 10. Percent of dispatching truck trials that were correctly dispatched, dispatched when not full, or overloaded.

Dispatching Trucks Performance Summary

Similar to the performance data for the Receiving Packages task, younger adults had a greater percentage correct compared to older adults for the Dispatching Trucks task. Among the two types of errors that could be committed, older adults had a higher rate of overloaded trucks compared to younger adults, whereas the percentage of not full trucks did not differ significantly between the two age groups. Additionally, younger adults experienced an effect of workload, such that high workload was associated with decreased performance compared to the low workload group. This pattern was not observed among older adults.

To gain a deeper understanding of what may be driving these performance differences in the Dispatching Trucks task, an in-depth analysis of how participants used the automation to accomplish the task was necessary. Automation use will be described in terms of compliance, reliance, and truck views.

Compliance

Compliance refers to the state in which the automation presents an alert or instruction to the participant. By complying, participants did not view the truck to verify that the alert was correct. If participants did view the truck to confirm that the automation's directive was correct, that would be considered an act of non-compliance. By examining the compliance data, it is possible to understand how participants reacted in the Dispatching Trucks task when they were presented with a specific instruction from the system to dispatch a truck.

Compliance was calculated as the percentage of trucks that were not viewed during the alarm state. The automated system was set to be 70% reliable, with 15% of the remaining trucks false alarm trials and the last 15% miss trials. Because the construct of compliance requires that an alert be present, even if it is an early alert (which would be the case with the false alarm trials), this excludes miss trials from the compliance analysis, as the miss trials do not contain an alert at any point in time. Therefore, compliance was examined in 136 of the total 160 truck trials.

Compliance was examined separately for trials in which the automation was correct and trials in which the automation was incorrect. Of the 136 trials in which compliance was examined, 112 were correct trials and 24 were incorrect trials. The automation incorrect trials included the false alarm trials only, rather than the false alarm and miss trials, because as previously mentioned, miss trials do not contain a compliance state because an alert is never presented. In an attempt to understand how participants' compliance behavior compared to optimal compliance behavior, optimal compliance was defined as complying in 100% automation correct trials and complying in none or 0% of the automation incorrect trials.

Figure 11 and Figure 12 depict compliance as a function of workload for younger and older adults, respectively. Although the omnibus ANOVA did not reveal a main effect of workload on compliance during automation correct or incorrect trials, planned contrasts showed significant differences between the workload groups for younger adults. When the automation was correct as well as incorrect, younger adults' compliance increased as the workload increased. That is, they were more likely to refrain from viewing the truck when workload was high. Specifically, planned contrasts revealed that younger adults in the high workload group complied more than younger adults in the low workload group in trials where the automation was correct, as well as trials where the automation was incorrect, $t(39) = -2.10$, $p = .04$, $t(39) = -2.35$, $p = .02$. No significant differences were found between younger adults in the low and moderate or moderate and high workload groups for automation correct or incorrect trials all ps > .21.

Note that when the automation is incorrect, younger adults in the high workload groups were more likely to comply with the automation, meaning that they did not verify its instruction before sending the truck, leading to a situation in which they erroneously dispatch a truck when it was not full. Recall that the Dispatching Trucks performance data revealed younger adults in the high workload group had significantly more instances of dispatching trucks that were not full. This finding is likely a result of the high group's greater compliance with the automation when it was committing a false alarm.

Among the older adults, there were also no significant differences between any of the workload groups for either automation correct or incorrect trials, all $p > .21$ (see

Figure 12). Additionally, the workload by age interaction was not statistically significant for compliance when the automation was correct or incorrect, all *p*s > .45.

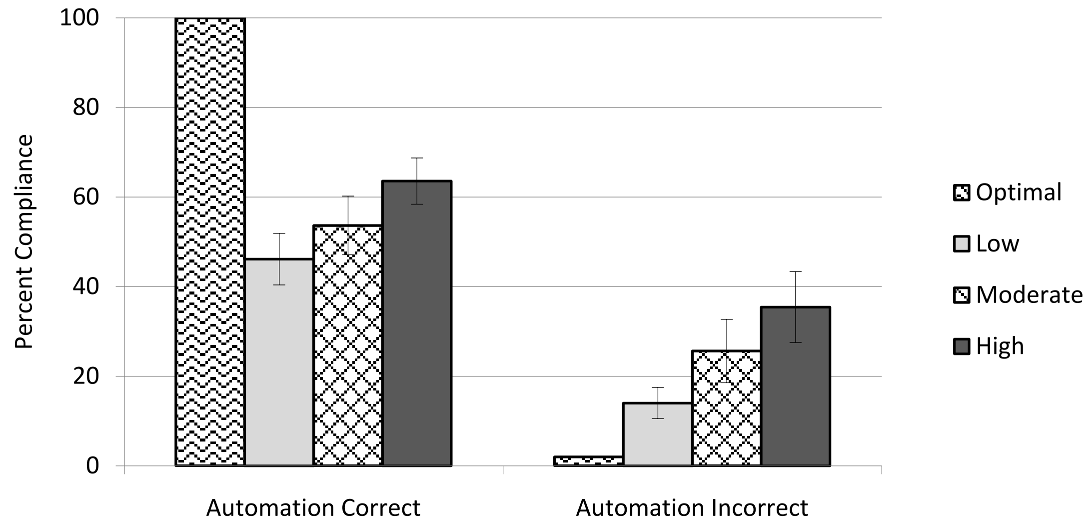


Figure 11. Younger adults' compliance, divided by workload groups. Optimal compliance is 100% for correct trials and approximately 0% for incorrect trials.

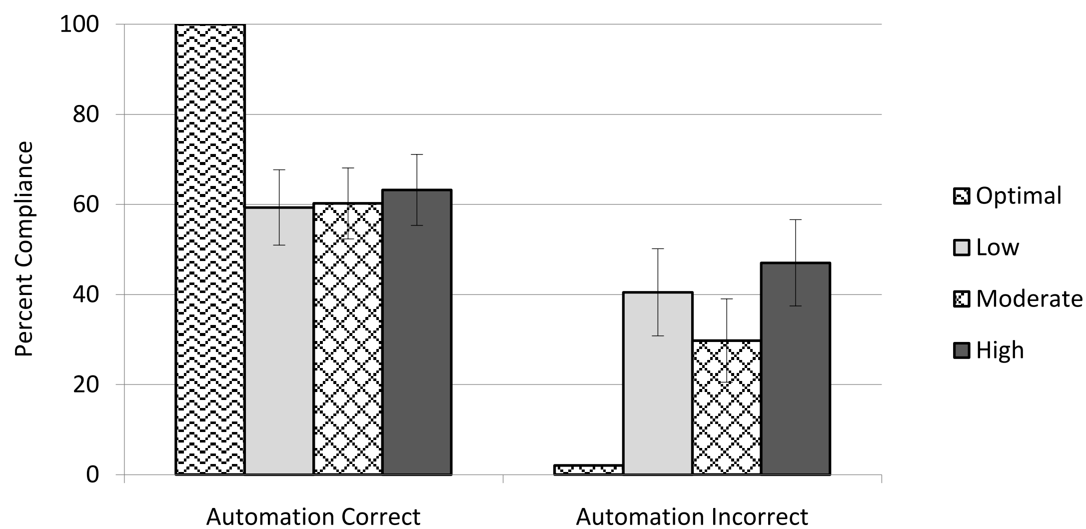


Figure 12. Older adults' compliance compared to optimal, divided by workload groups. Optimal compliance is 100% for correct trials and approximately 0% for incorrect trials.

In addition to these workload differences, age-related differences were also observed (see Figure 13). Older adults complied significantly more than younger adults when the automation was incorrect ($F(1, 78) = 4.08$, $p = .05$, $\eta^2 = .05$), and although the pattern was consistent when the automation was correct as well, the age-related difference was not statistically significant. Because older adults were complying with the automation when it erred at a greater rate than younger adults, this means that they were more likely to not detect the false alarm and inadvertently send a not full truck. However, the Dispatching Trucks data presented earlier did not display a significant age-related difference in the percentage of not full trucks, as would be expected. This may be the case for a number of possible reasons.

First, when the incorrect alert (false alarm) is presented, it may simply be the case that older adults do not notice the alert at all, which is why they do not check the truck and thereby comply with the automation. This would also mean that they fail to dispatch the truck at all, which would lead to an overloaded truck, and the Dispatching Trucks data did show that older adults had significantly more overloaded trucks than younger adults. A second potential explanation may be that older adults do notice the alert and plan to view the truck and/or dispatch the truck at some point, but fail to do so before the truck overloads. Lastly, older adults may notice the alert from the automation but choose not to look at the truck and therefore comply because the cost of non-compliance is particularly high. This point will be revisited later.

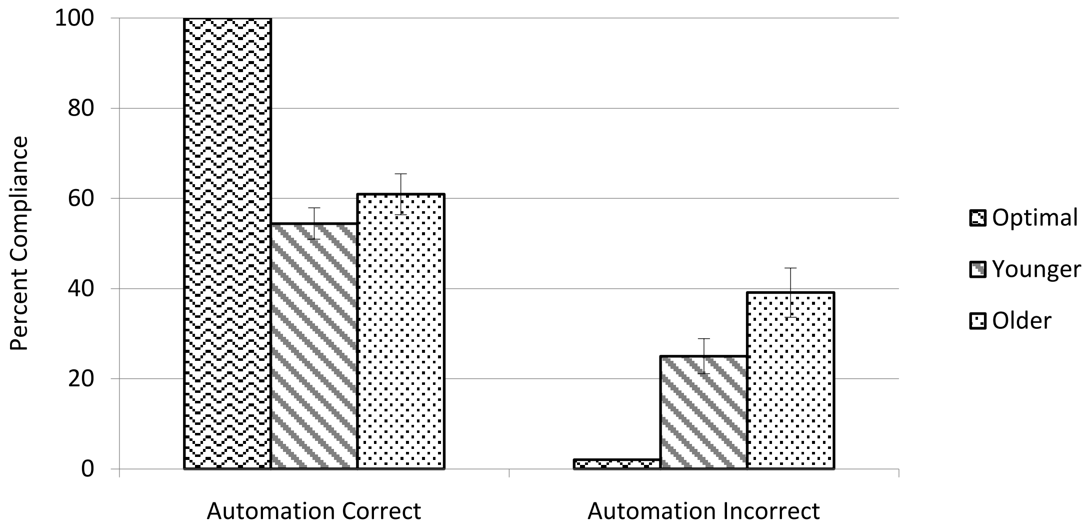


Figure 13. Compliance for younger and older adults. Optimal compliance is 100% for correct trials and approximately 0% for incorrect trials.

Another pattern to note, as depicted in Figure 13, is that compliance is lower in trials where the automation is incorrect, compared to the automation correct trials. This is true for both younger and older adults. Compliance behavior differs as a function of whether the automation is providing accurate or inaccurate information. This may indicate that participants have some ability to recognize when the automation is erring, and adjust their compliance behavior accordingly. The only cue available to participants that could be used to identify automation false alarms is the temporal nature of this task. Although the time required for the truck to fill varied between 12 and 22 seconds, it may still be feasible that participants used some kind of time-based strategy to determine whether an alert was early or not.

Reliance

Reliance refers to use of the automation during the silent state, which was any time there was not an alert from the automation present. By relying, participants did not view the truck during the time that the automation was silent. By examining the reliance data it is possible to gain further insights into participants' use of the automation and the resulting effect this had on performance in the Dispatching Trucks task.

Reliance was calculated as the percentage of trucks that were not viewed during the silent state. Reliance was examined separately for trials in which the automation was correct and trials in which the automation was incorrect. Of the 160 trials in which reliance was examined, 112 were correct trials and 48 were incorrect trials. The automation incorrect trials included both misses and false alarms. Similar to the compliance data, reliance behavior was compared to optimal reliance behavior. Optimal reliance was defined as complying in 100% automation correct trials and relying in none or 0% of the automation incorrect trials.

Younger adults' reliance was not significantly affected by workload, and this was true for reliance when the automation was correct as well as incorrect (see Figure 14). Planned contrasts did not reveal significant differences between any of the workload groups for younger adults, all *p*s > .42. Numerically, younger adults relied on the automation less as the workload increased, opposite of the pattern observed for compliance behavior.

Older adults' reliance behavior more closely mimics the patterns observed in the compliance data (see Figure 15). As workload increased, reliance on the automation increased, and this was true when the automation was correct as well as incorrect. Planned contrasts revealed that older adults in the high workload group relied significantly more than the moderate workload group both when the automation was correct and incorrect, $t(39) = -2.15$, $p = .04$, $t(39) = -2.37$, $p = .02$. There was not a significant age by workload interaction for automation correct trials or automation incorrect trials, all *ps* > .46.

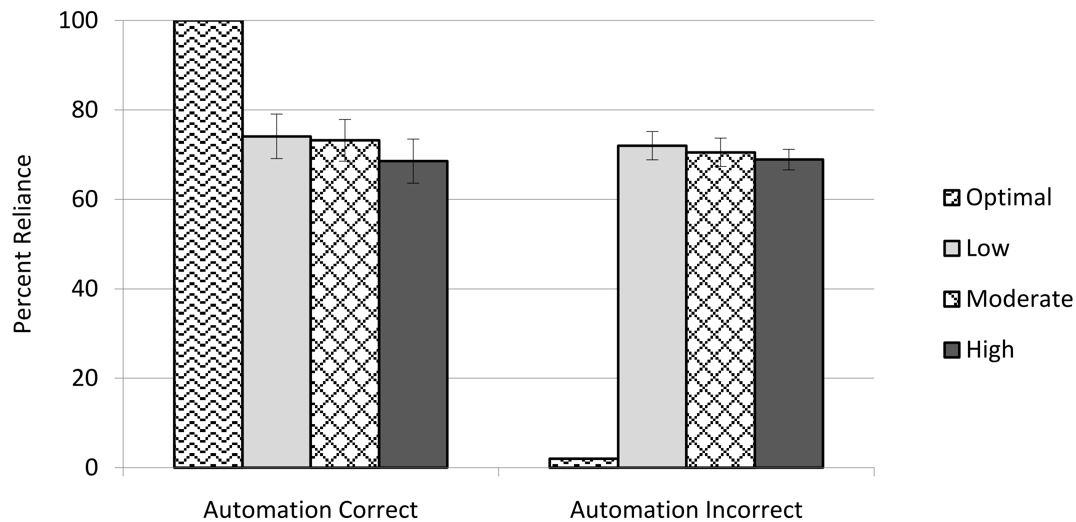


Figure 14. Younger adults' reliance, divided by workload groups. Optimal reliance is 100% for correct trials and approximately 0% for incorrect trials.

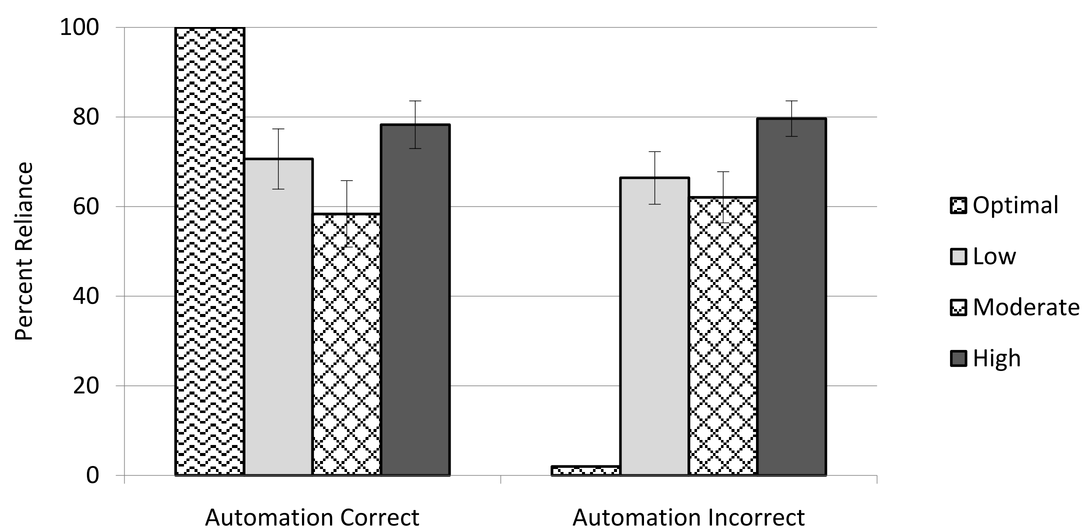


Figure 15. Older adults' reliance, divided by workload groups. Optimal reliance is 100% for correct trials and approximately 0% for incorrect trials.

Contrary to the patterns that emerged in the compliance data, age-related differences were not observed for reliance behavior (see Figure 16). No main effect of age was found for reliance during automation correct of incorrect trials, all *ps* > .54.

Both younger and older adults relied on the automation for approximately 70% of the truck trials, regardless of whether the automation was correct or incorrect. In addition to the lack of an age-related difference, the reliance data also fail to exhibit any indication that participants adjusted their reliance behavior when the automation erred. Contrast this finding to the pattern observed in the compliance data wherein participants reduced their compliance by almost half when the automation was incorrect. Because this pattern was not found for reliance, this suggests that the errors that affect reliance, namely misses, are more difficult to recognize than false alarms.

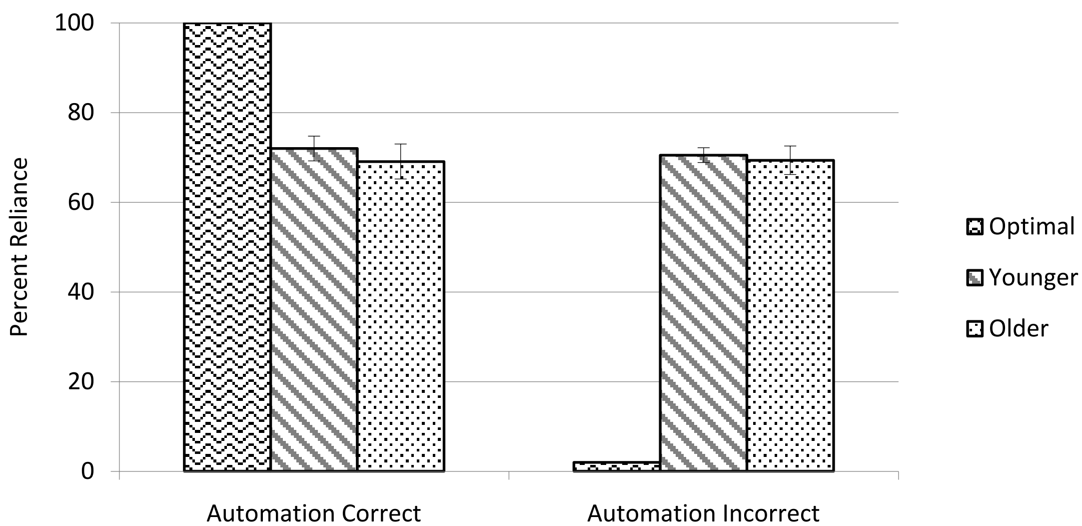


Figure 16. Reliance for younger and older adults. Optimal reliance is 100% for correct trials and approximately 0% for incorrect trials.

Summary of Compliance and Reliance Data

The analyses of compliance and reliance provided insights regarding how the workload manipulation in the Receiving Packages task influenced the way participants used the automation. As workload increased, younger adults were more likely to comply with the automation, both when the automation was correct and incorrect. Because younger adults in the high workload group complied with the automation more than the low workload group, the high workload group had a significantly higher rate of trucks that were dispatched before they were full. However, their reliance behavior was not significantly influenced by workload. Conversely, older adults' reliance increased as workload increased, whereas their compliance was not significantly affected by workload.

Further, the data revealed that older adults complied with the automation more than younger adults, but only when the automation was incorrect. Additionally, the compliance data showed that participants reduced their compliance for trials in which the automation erred, suggesting they were able, to some degree, to recognize when the automation was committing a false alarm. Younger and older adults relied approximately the same amount, and neither group reduced their reliance for instances where the automation erred, suggesting that misses may be harder to detect.

Truck View Details

In addition to examining compliance and reliance, a more detailed inspection of the instances in which participants chose not to comply or rely was performed. In the previous analyses, trials in which participants viewed the truck once, twice, or more were all scored the same, as an act of non-compliance or non-reliance. The actual number of times a particular truck was viewed was not factored into the analysis. In the following analyses, the total number of truck views will be considered. Additionally, the mean number of times participants viewed a single truck was examined, as well as the mean duration of each truck view.

Total Truck Views

In addition to examining whether a participant viewed a truck during the silent start or alarm state (non-relied or non-complied, respectively), the total number of views across the entire experiment was also examined.

Figure 17 and Figure 18 depict truck views for younger and older adults. Workload did not significantly influence the number of views across all trials or for the subset of automation correct trials, all *p*s > .29. However, planned contrasts revealed that among younger adults, the low workload group had significantly more views compared to the high workload group when the automation was incorrect, $t(39) = 2.05$, $p = .05$. No significant differences were found among older adults, all *p*s > .15.Although significant differences were not consistently found, the pattern suggests that as workload increases, the frequency of views tends to decrease, at least among younger adults.

Although younger and older adults did not differ in the number of views overall or when the automation was correct, younger adults had significantly more views than older adults during trials in which the automation erred, $F(1,78) = 9.28$, $p < .01$, $\eta^2 = .10$. This may have contributed to younger adults' superior performance in Dispatching Trucks.

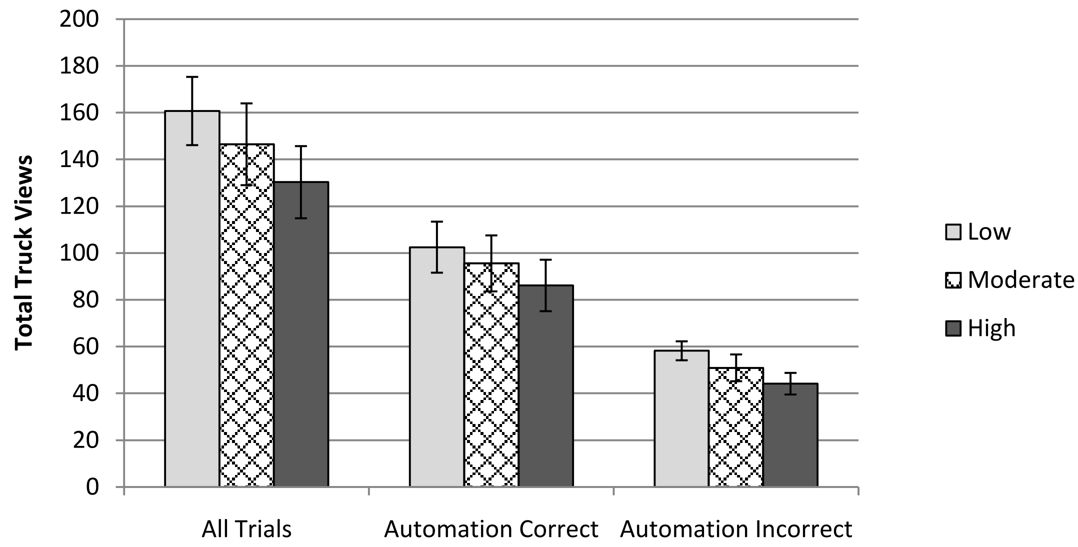


Figure 17. Younger adults' total truck views for all trials, automation correct trials, and automation incorrect trials.

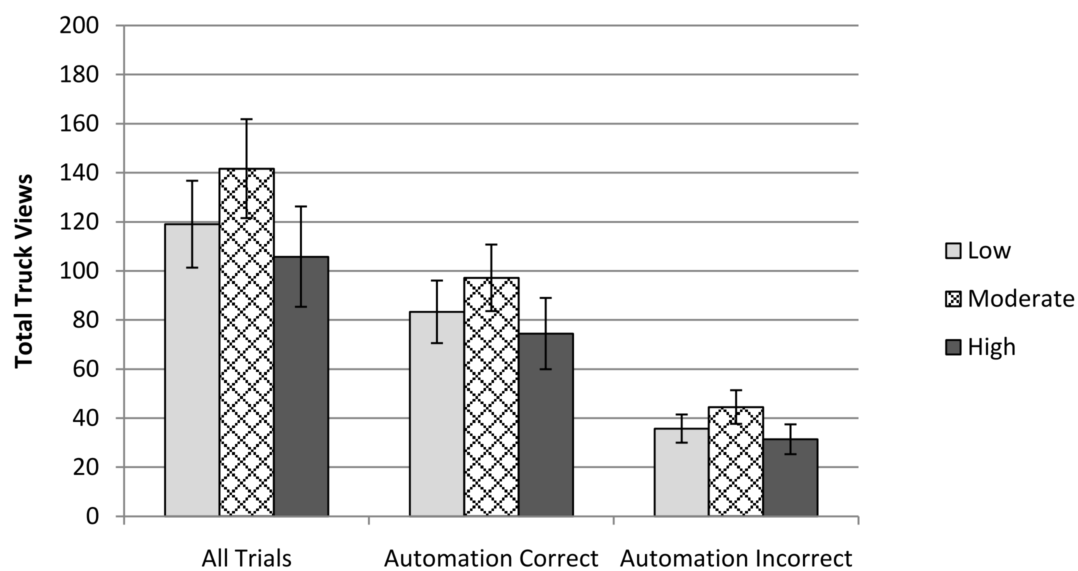


Figure 18. Older adults' total truck views for all trials, automation correct trials, and automation incorrect trials.

Views Per Truck

Because it was possible to view each truck more than once, the average number of views per truck was also examined (see Figure 19). The analysis did not reveal a main effect of age or workload group, or a significant interaction, all *p*s > .09. Planned contrasts did not yield any significant differences between workload groups for either age group, all *p*s > .23. Although no significant patterns were detected, note that for all groups, the average number of views per truck was above one, suggesting that on average, participants were looking at the trucks once or twice, rather than once or not at all.

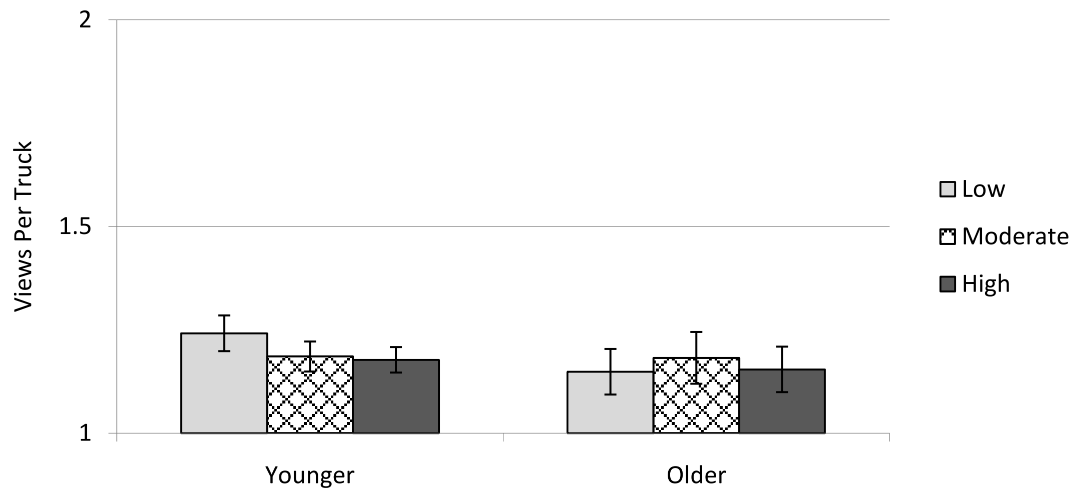


Figure 19. Views per truck by age and workload group.

Duration of Truck Views

Another aspect of the truck views that was of interest was the length of time or duration of the truck views (see Figure 20). Recall that when participants viewed the truck, they were taking time away from the primary task of Receiving Packages to verify or double-check the automation. On average, older adults' truck views were significantly longer than younger adults', $F(1,69) = 39.38$, $p < .01$, $\eta^2 = .25$. There was not a main effect of workload group, $p = .60$, or an interaction between age and workload group, $p = .35$. Planned contrasts did not reveal significant differences between the workload groups for either younger or older adults, all *p*s > .10.

When older adults viewed the truck they spent a much longer time doing so before switching back and working on Receiving Packages compared to younger adults, suggesting perhaps that they required a longer period of time to process the information presented and potentially project how long it would be before the truck was actually filled.

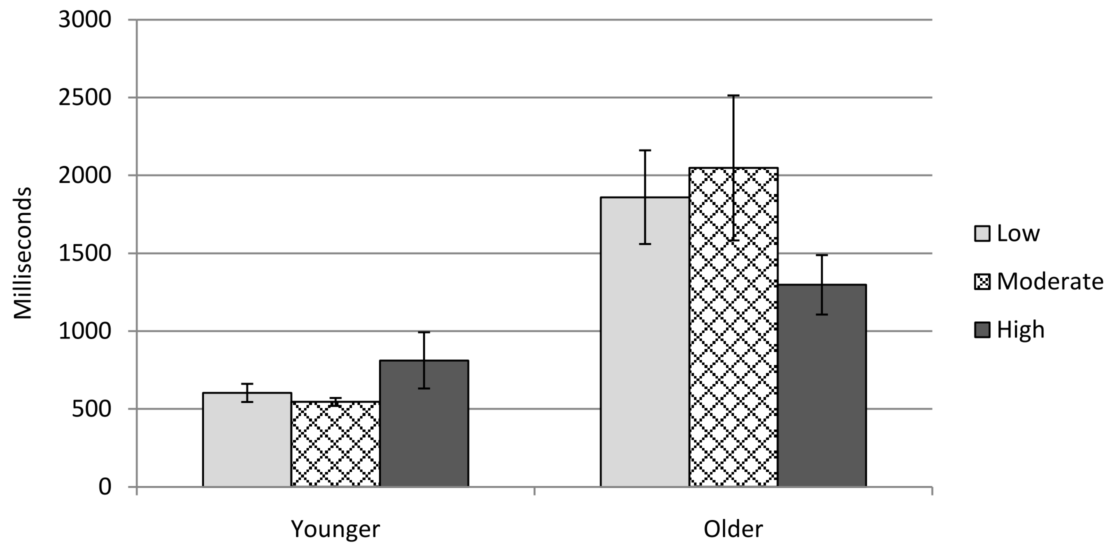


Figure 20. The average length of time spent looking at a truck.

Truck View Details Summary

An in-depth analysis of truck viewing behavior revealed that the total number of truck views was not significantly affected by workload or age. However, when the total number of truck views was examined in the context of automation accuracy, younger adults had significantly more truck views than older adults when the automation was incorrect. Views per truck did not vary by workload or age group. However, when the mean length of a truck view was examined, older adults were found to have significantly longer truck views than younger adults.

Subjective Measures

Various measures were administered over the course of the experiment to assess the participants' subjective experience relating to use the automated system, including a rating of trust and perceived reliability of the system. These two measures have been found to be predictive of compliance and reliance in previous research. Additionally, a modified version of the NASA-TLX was used to evaluate the participants' workload as it related to the overall task of managing the warehouse, not to either of the two tasks specifically.

Trust

The trust rating data is presented in Figure 21. The analysis revealed no main effect of workload, $p = .24$ and planned contrasts similarly revealed no differences between workload groups for either younger or older adults, all ps > .178. However, there was a main effect of age, such that older adults rated their trust in the automated aid higher than younger adults, $F(1, 78) = 5.06$, $p = .03$, $\eta^2 = .06$.

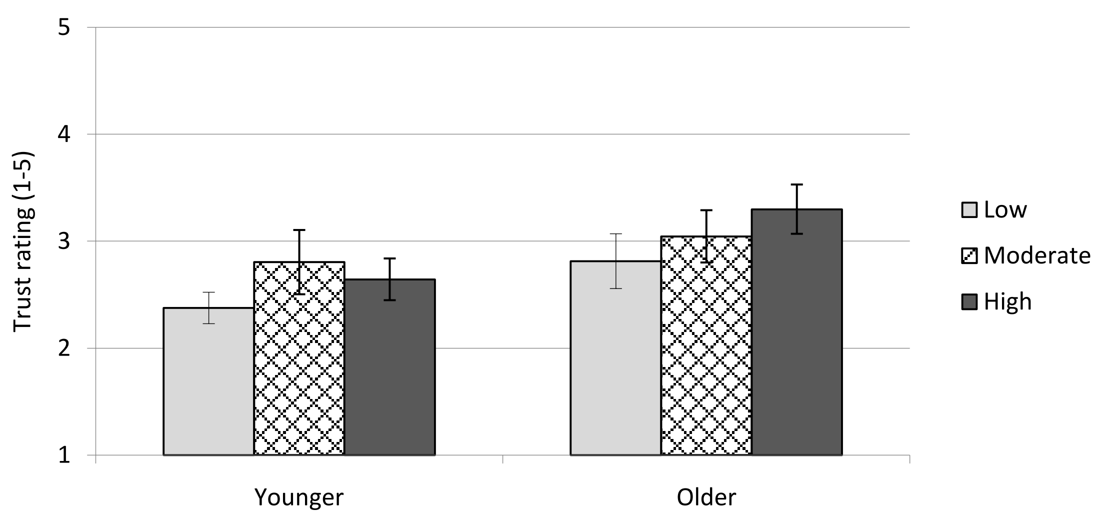


Figure 21. Trust by age and workload group. The trust rating ranged from 1 – Very Little Trust, to 5 – Very Much Trust.

Perception of Reliability

A Pearson correlation analysis revealed a significant and positive correlation between trust rating and perception of reliability, $r(82) = .47$, $p < .01$. This suggests that trust in the automation and the perception of the automation's reliability are positively associated. However, although a main effect of age was observed for trust, no such effect was observed for perception of reliability, $p = .22$ (see Figure 22). Additionally, there was not an effect of workload group or an interaction between age and workload group, all *p*s > .23. Planned contrasts between the workload groups for younger and older adults did not reveal any significant differences for average perceived reliability, all *p*s > .07. One interesting finding from these data was that both younger and older adults were highly accurate in terms of their perceptions of how reliable the system was. Recall actual system reliability was 70%.

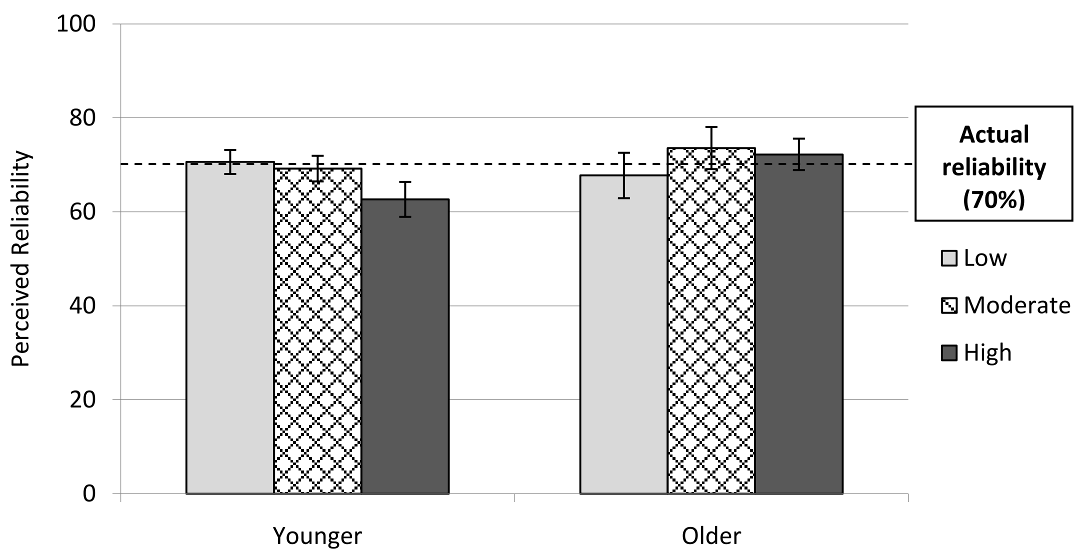


Figure 22. Perceived reliability by age and workload group. The dotted line represents the actual reliability of the automation.

NASA-TLX

The NASA-TLX was administered to participants to assess their overall level of workload. This assessment was therefore not tied to a particular task, but to the high-level task of managing the warehouse. Figure 23 displays the NASA-TLX scores for both older and younger adults by workload group. The analysis revealed a significant effect of workload group, $F(2, 77) = 3.58$, $p = .03$, $\eta^2 = .07$. Planned contrasts revealed younger adults in the low workload group reported a significantly lower average NASA-TLX score that younger adults in the high workload group, $t(39) = -2.2$, $p = .03$. There was not a significant difference between younger adults in the low and moderate workload groups or the moderate and high workload groups, all $p > .13$.

Older adults in the moderate workload group reported a significantly lower average NASA-TLX score than older adults in the high workload group, $t(39) = -2.44$, $p = .02$ (see Figure 23). There was not a significant difference between older adults in the low and moderate workload groups, $p = .25$, or the low and high workload groups, $p = .21$. A main effect of age on NASA-TLX scores was also found, $F(1, 77) = 9.49$, $p < .01$, $\eta^2 = .10$ showing that older adults typically reported higher NASA-TLX ratings than younger adults.

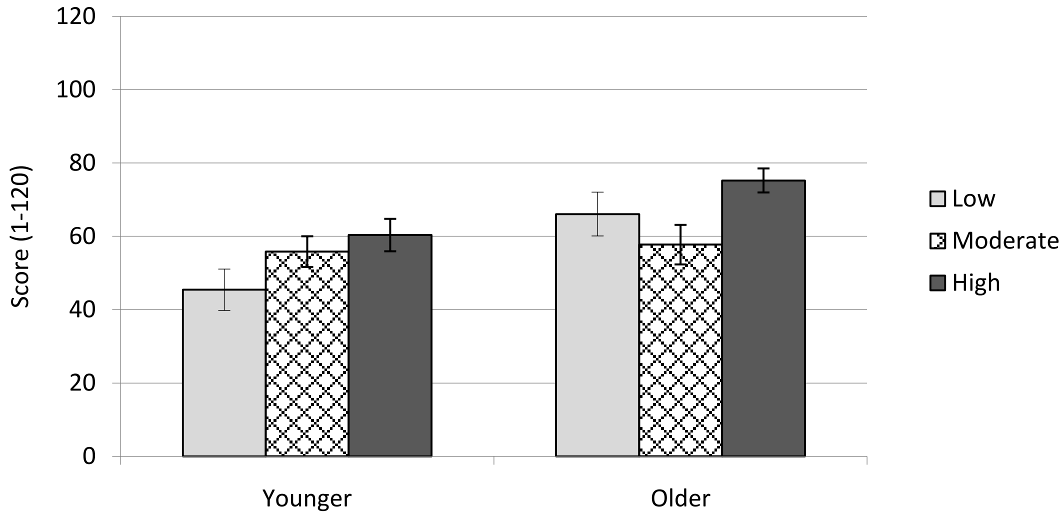


Figure 23. NASA-TLX score by age and workload group.

The NASA-TLX score is a composite measure comprising six sub-scales. Table 4 displays the composite as well as sub-scale scores for each of the six between-subject groups. Mental demand, temporal demand and effort appeared to be the most taxing aspects of the task for both younger and older adults.

Table 4

Mean NASA-TLX Composite and Sub-scale Scores

	Composite	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration
Younger adults							
Low workload	45.43	8.79	4.13	9.84	6.50	9.29	6.96
Moderate workload	55.84	11.23	3.78	13.71	5.80	12.71	8.71
High workload	60.35	10.74	7.14	13.08	8.74	12.60	8.26
Older adults							
Low workload	66.09	13.37	9.43	13.30	7.97	13.38	8.57
Moderate workload	57.74	11.85	7.11	12.24	6.51	11.80	8.46
High workload	75.26	15.05	9.36	15.19	11.00	15.05	9.46

In addition to the NASA-TLX, the subjective workload assessment technique (SWAT) was also used to gather information regarding workload. These data revealed an age-related difference, such that older adults reported higher SWAT scores than younger adults, $F(1, 77) = 17.12$, $p < .01$, $\eta^2 = .16$. Neither a main effect of workload nor an age by workload interaction were not found, $p > .11$. These data can be found in Appendix F.

Subjective Measures Summary

Older adults displayed higher levels of reported trust in the automated system compared to younger adults, however there were no significant effects of workload on trust. Trust was positively correlated with perception of reliability, although there was not an effect of workload or age on perceived reliability. Lastly, the high workload groups for older and younger adults reported higher scores on the NASA-TLX. Additionally, older adults reported greater workload on the NASA-TLX compared to younger adults.

Key Findings

Workload experienced in the Receiving Packages task had an effect on a number of factors. As workload increased, performance on the Receiving Packages was affected for both younger and older adults. Although younger adults also experienced an effect of workload on Dispatching Trucks performance, older adults did not. The compliance data showed that as workload increased, compliance with the automation's instructions increased, but only for younger adults. Reliance was similarly affected by workload, but for older adults only. Lastly, as workload in the Receiving Packages task increased, so too did perceived workload as measured by the NASA-TLX.

Significant differences between younger and older adults were found in a number of areas. Younger adults outperformed older adults on both tasks. When the automation was incorrect, younger adults complied with it to a lesser extent and also had fewer total truck views than older adults. This suggests that younger adults may be better able to identify when the automation is committing a false alarm, resulting in a change in their compliance behavior. Older and younger adults did not differ in terms of their reliance on the automation. When older adults viewed the truck they did so for a longer time than younger adults. Older adults also reported greater trust in the automation and higher workload.

CHAPTER 4

CONCLUSION

This study was designed to assess whether workload affects the manner in which a person uses an automated system, and to compare younger and older adults' automation interactions. Increasing the workload imposed in the Receiving Packages task led to reduced performance accuracy in the Receiving Packages task for both age groups, but the Dispatching Trucks performance accuracy was reduced for younger adults only. Younger adults correctly matched more barcodes than older adults in the Receiving Packages task and also correctly dispatched more trucks than older adults in the Dispatching Trucks task. An analysis of the different errors participants could have made in each task revealed that older adults incorrectly matched and timed out on more barcodes than younger adults in the Receiving Packages task. Within the Dispatching trucks task, older adults overloaded more trucks than younger adults, but the two groups did not significantly differ in the percent of trucks that were dispatched early.

To understand how participants' use of the automation led to the observed patterns of performance, participants' compliance with and reliance on the automation was examined. More specifically, compliance and reliance were both examined separately for the subset of trials in which the automation was correct and the trials in which the automation was incorrect. Global measures of compliance or reliance make it difficult to compare whether one group's compliance or reliance is more appropriate than another's, because this depends on whether the automation is correct or incorrect. When the automation is correct, high compliance and reliance is optimal, because viewing the truck would be an ineffective use of time as the automation is providing accurate information. When the automation is incorrect, compliance and reliance should be low, because viewing the truck is required to correct the automation's error and dispatch the truck successfully.

Among younger adults, as workload increased, compliance increased, regardless of whether the automation was correct or incorrect. More specifically, the high workload group had higher compliance than the low workload group when the automation was incorrect (committing a false alarm). Because the high workload group complied with the automation when it was giving the alert to dispatch early, that group dispatched significantly more trucks that were not yet full than the low workload group. Younger adults' reliance behavior did not follow a similar pattern. Higher workload did not lead to higher reliance; there were no significant differences between the workload groups.

The compliance and reliance behavior of older adults was markedly different than that of younger adults. Load did not have a significant effect on compliance for older adults. However, the pattern of compliance differed between automation correct and incorrect trials. Specifically, when the automation was correct, the workload groups complied approximately the same amount, but when the automation was incorrect, the high workload group complied numerically more than the moderate group, although this difference was not statistically significant. Reliance on the automation was higher among the high workload group compared to the low workload group, both when the automation was correct and incorrect.

Comparing younger and older adults' use of the automation revealed that older adults complied more with the automation than younger adults, but this difference was significant only when the automation was incorrect. Older adults' greater compliance with the automation when it was committing an error likely led to the performance discrepancy between younger and older adults in the Dispatching Trucks task.

The analysis of compliance and reliance as a function of automation accuracy also revealed that participants reduced their compliance when the automation was incorrect compared to when it was correct. This lends credence to the notion that there is some level of error recognition happening among participants, at least for the false alarms. If this were not the case, then compliance should be as prevalent during the automation error trials as it is during the automation correct trials. The reliance data did not exhibit any sign of error recognition, as reliance did not vary from automation correct to incorrect trials. This pattern is in line with findings suggesting that false alarms may be easier to detect than misses (Johnson, 2004).

Examining the length of time participants spent viewing a truck revealed that older adults had significantly longer truck views than younger adults by approximately a factor of three. It is not clear whether this difference occurred by choice or by necessity, but regardless of the reason it means that viewing the truck carried a higher cost for older adults, because it meant more time away from the Receiving Packages task which may disrupt performance of that task. Therefore, older adults may comply and rely more than younger adults because the cost of checking the automation is higher for older adults. Indeed, Ezer's (2006) work showed that when the cost of verifying automation is high, participants will comply with the automation to a greater degree.

Further, there is considerable evidence present in the cognitive aging literature to suggest that the cost of verifying the automation is higher for older adults because of difficulties associated with dual-tasking and task-switching, such as higher switch costs (Cepeda, Kramer, & Gonzalez de Sather, 2001; Kramer & Madden, 2008). Consider the implications of these data if applied to a system such as an automated in-car navigation system. If the primary task is driving and the secondary is double-checking that automation, then spending an extended period of time away from the main task of driving could be potentially disastrous.

Higher workload may be associated with greater compliance and reliance because as workload increases, the cost of verifying the automation increases. That is, in high workload states, less resources or attention are available to devote to verifying the automation because the Receiving Packages task is more difficult and requires greater attention and effort compared to low workload states. Therefore, when switching to viewing the truck, it requires more time to processing the information because the participants must disengage more of their cognitive resources from the Receiving packages task. Although the difference was not significant, younger adults in the high workload group did look longer at the truck compared to the low and moderate workload groups.

Both age groups estimated the reliability of automation fairly accurately and workload did not significantly affect the perception of reliability. This serves as support against any claims that older adults relied or complied more because they thought the system was more reliable.

Error Detection

The data revealed that when the automation committed a false alarm, younger adults displayed lower compliance than older adults, although both age groups reduced their compliance for false alarm trials compared to correct trials. However, reliance did not vary between automation correct and incorrect trials for either age group, suggesting that automation misses may be more difficult to identify.

The compliance data suggest that there may be an age-related difference in error detection ability. Because the only information available to aid in error detection was temporal cues, the age-related difference might be due to age-related differences in time estimation. Although there is some debate in the literature, research has shown that significant age-related differences in judgments of time duration exist (Block, Zakay, & Hancock, 1998). If older adults cannot accurately determine how much time and has passed, this may influence their ability to determine whether an alert was early or whether enough time had passed for an alert to appear.

Practical Implications

For certain automation behaviors, such as compliance, participants were able to adjust their behavior upon recognizing that the automation was committing an error. Training users of automated systems to detect errors and modify their behavior accordingly is an important component of any training regimen. It would make very little sense to train users to comply at a particular level without regard to whether the automation was erring or not. Depending on the system in question this may be relatively easy or difficult to do. Certainly, in some automated systems it may be quite easy to detect when the automation is erring, whereas in others this may be particularly challenging or unfeasible due to the complexity of the system. However, understanding how to increase a user's ability to identify errors should be a training priority.

Additionally, designers of automated systems that may require verification when an instruction is presented must consider the associated costs of verification, as these may vary between younger and older users, and will likely have consequences for how individual comply with and rely on the automation.

Next Steps

Investigating how users of automated systems are able to detect automation errors is an important direction for future research. This study demonstrated that participants seemed to be able to use subtle, temporal cues to identify instances in which the automation committed a false alarm. However, it is not clear what other types of cues or factors may contribute to error identification. Wilkison (2008) showed that participants with accurate mental models were more likely to avoid using the automation's suggestion when it was inaccurate. Related factors such as experience and frequency of errors may allow users to better diagnose the automation as providing accurate information or faulty information, however these and other factors remain to be investigated.

APPENDIX A

GENERAL TRUST IN AUTOMATION

An automated system is a technologically-based system used to partially or fully assist the human in tasks involving sensing, detecting, information processing, making decisions and/or executing actions. Examples include automated teller machines (ATMs) and in-vehicle navigation systems.

Please circle the number that best describes your feeling or impression.

1. Automated systems are deceptive.

1 2 3 4 5 6 7

Not at All **Extremely**

2. Automated systems behave in an underhanded manner.

1 2 3 4 5 6 7

Not at All **Extremely**

3. I am suspicious of automated systems' intent, action, or outputs.

1 2 3 4 5 6 7

Not at All **Extremely**

4. I am wary of automated systems.

1 2 3 4 5 6 7

Not at All **Extremely**

5. Automated systems' actions have a harmful or injurious outcome.

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

6. I am confident in automated systems.

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

7. Automated systems provide security.

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

8. Automated systems have integrity.

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

9. Automated systems are dependable.

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

10. Automated systems are reliable.

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

11. I can trust automated systems.

1	2	3	4	5	6	7

Not at All **Extremely**

12. I am familiar with automated systems.

1	2	3	4	5	6	7

Not at All **Extremely**

13. To what extent do you think you could count on an Automated System to do its job?

1	2	3	4	5	6	7

Not at All **Completely**

14. Overall, how much would you trust an Automated System?

1	2	3	4	5	6	7

Not at All **Completely**

15. Please indicate how often you think an Automated System would provide correct information (using a %).

(Example: I think an Automated System would provide correct information ##% of the time)

_______%

APPENDIX B

INTERIM QUESTIONNAIRE

1) Please indicate, using a percentage, how often you thought the automated system alerted you at the correct time to dispatch a full truck (0-100%)?

    ____________%

2) How much did you trust the automated system to correctly alert you when a truck was full and ready to be dispatched?

1	2	3	4	5
Very Little		**Neutral**		**Very Much**

3) What percentage of the time did you view the truck when the automated system gave you an alert to dispatch the truck (0-100%)?

    ____________%

4) What percentage of the time did you view the truck when there was no alert present from the automated system (0-100%)?

    ____________%

Please answer the following questions by placing a mark in the appropriate box like this |X|.

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

APPENDIX C

PERFORMANCE AND STRATEGY QUESTIONNAIRE

1) Overall, how often you thought the automated system correctly alerted you to dispatch a full truck (0-100%)?

________%

2) Overall, how much did you trust the automated system to correctly alert you when a truck was full and ready to be dispatched?

1	2	3	4	5
Very Little		**Neutral**		**Very Much**

3) Overall, what percentage of the time did you depend solely on the automated system to determine when a truck was full and ready to be dispatched?

_________%

4) Have you ever worked in a Warehouse loading and shipping department?

No Yes
If Yes, how long did you work there?
______ years

5) Have you ever done anything similar to the tasks you performed over the past two sessions?

No Yes
If Yes, please describe it below.

____
____
____
____

6) How challenging was the overall warehouse manager task?

1	2	3	4	5	6	7
Very Easy			**Neutral**			**Very Difficult**

7) Did you pay attention to one task more than the other?

1	2	3	4	5	6	7
Mostly to Dispatching Trucks			**Both Tasks Equally**			**Mostly to Receiving Packages**

8) Did you understand how to Receive Packages (that was the task where you had to match the barcode)?

No

Yes

If Yes, please explain how you approached the task on Day 1.

____

____

____

____

____

If Yes, did your approach or method change on Day 2? Please explain.

____

____

____

____

____

9) How challenging was the Receiving Packages task on Day 1?

1	2	3	4	5	6	7
Very Easy			**Neutral**			**Very Difficult**

10) How challenging was the Receiving Packages task on Day 2?

1	2	3	4	5	6	7
Very Easy			**Neutral**			**Very Difficult**

11) Did you understand how to Dispatch Trucks (that was the task where the automated system told you when to dispatch the truck)?

No Yes

If Yes, please explain how you approached the task on Day 1.

If Yes, did your approach or method change on Day 2? Please explain.

12) How challenging was the Dispatching Trucks task on Day 1?

1	2	3	4	5	6	7
Very Easy			**Neutral**			**Very Difficult**

13) How challenging was the Dispatching Trucks task on Day 2?

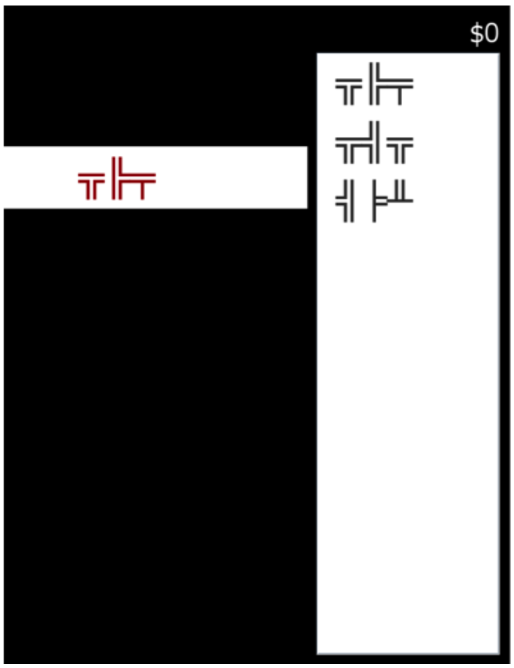
1	2	3	4	5	6	7
Very Easy			**Neutral**			**Very Difficult**

APPENDIX D

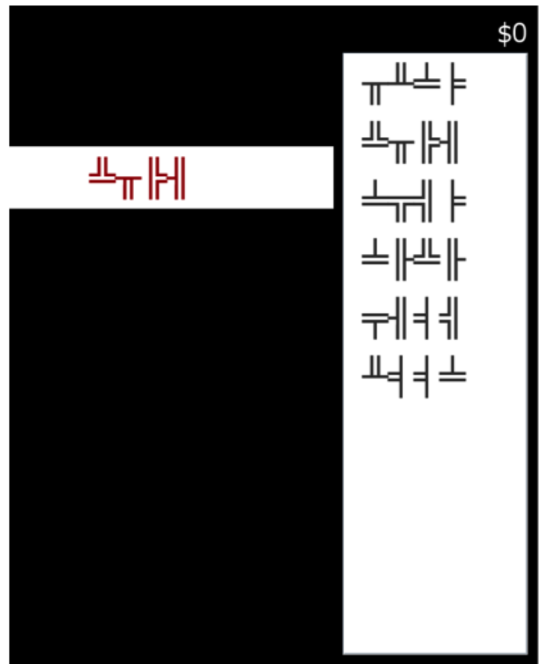
WORKLOAD MANIPULATION

Younger adults

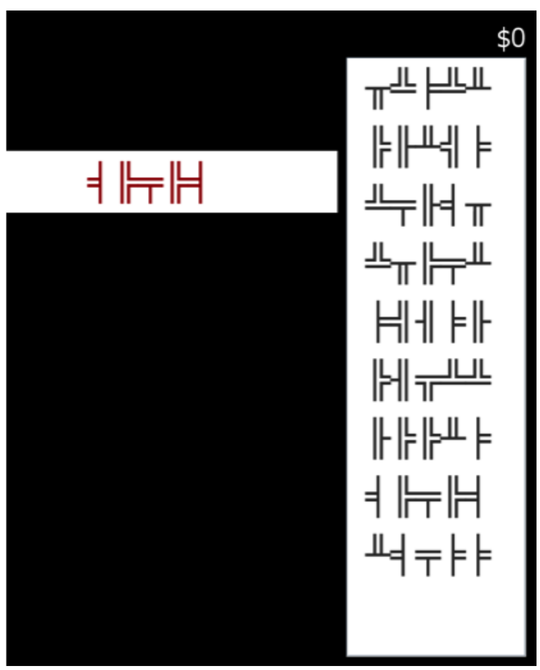
Low workload



Moderate workload

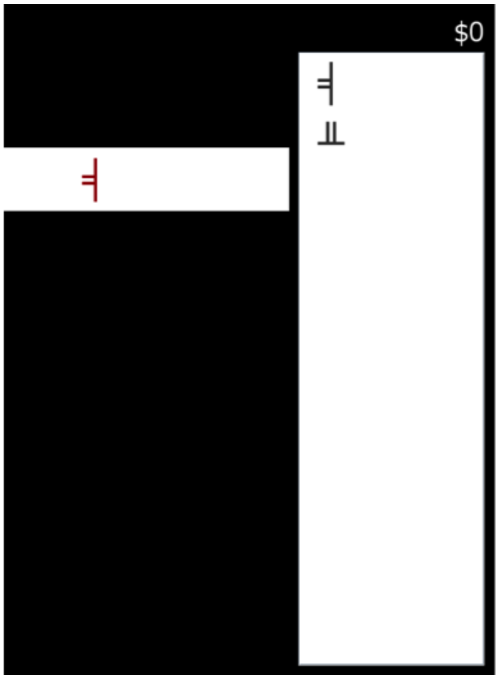


High workload

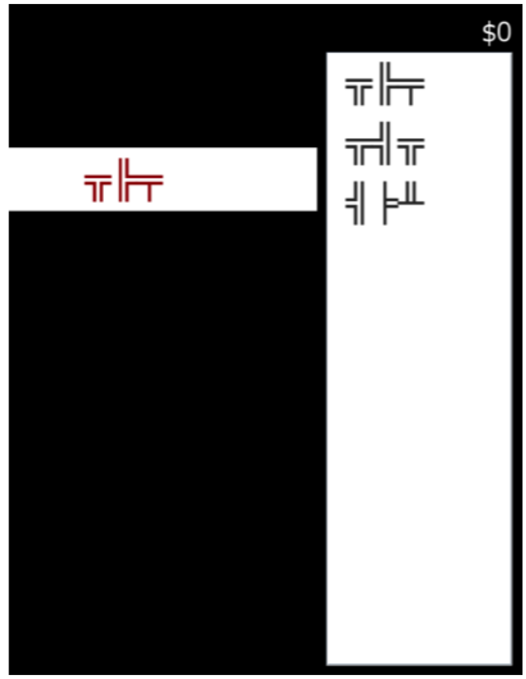


Older adults

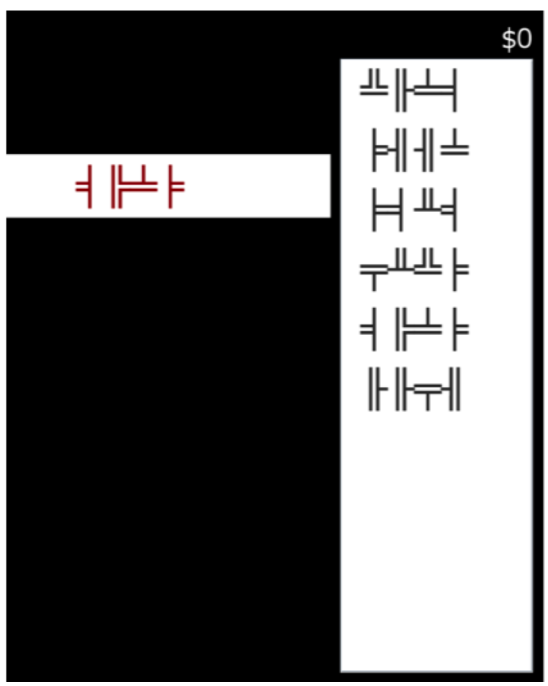
Low workload



Moderate workload



High workload



APPENDIX E

BLOCK LEVEL DATA

Table E1

Receiving Packages - Percent Correct: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	0.074	1	0.074	10.581	0.002
Workload	0.578	2	0.289	41.084	0.000
Age*Workload	0.021	2	0.010	1.475	0.235
Block	0.099	1.958	0.051	34.463	0.000
Block*Age	0.017	1.958	0.009	5.907	0.004
Block*Workload	0.040	3.916	0.010	7.008	0.000
Block*Age*Workload	0.003	3.916	0.001	0.571	0.680

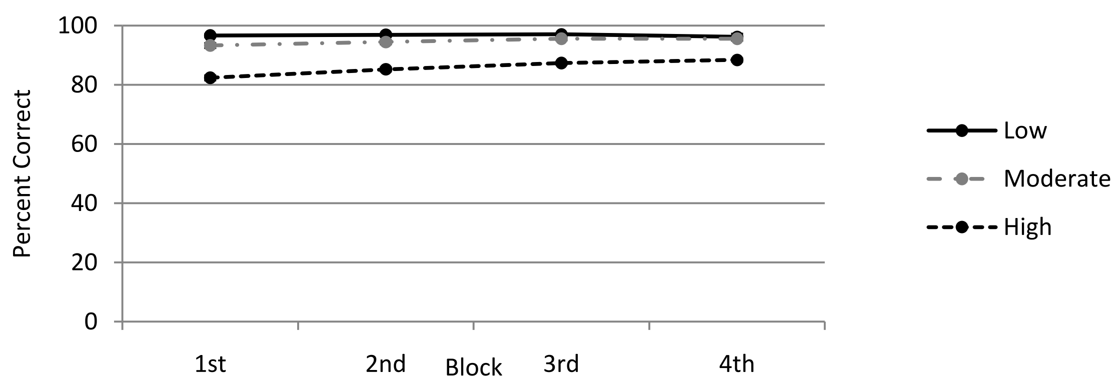


Figure E1. Younger adults' receiving packages performance across blocks.

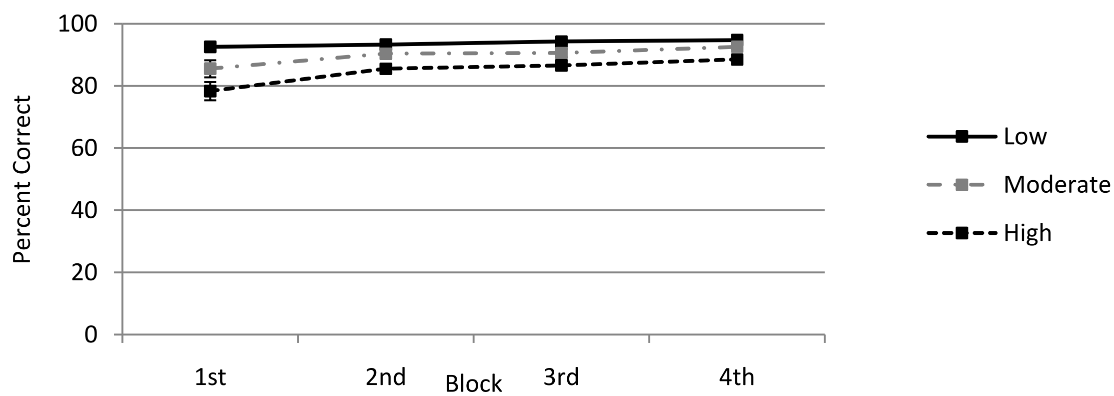


Figure E2. Older adults' receiving packages performance across blocks.

Table E2

Dispatching Trucks - Percent Correct: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	0.575	1	0.575	14.281	0.000
Workload	0.173	2	0.087	2.150	0.123
Age*Workload	0.056	2	0.028	0.696	0.501
Block	0.042	3	0.014	3.087	0.028
Block*Age	0.009	3	0.003	0.681	0.565
Block*Workload	0.007	6	0.001	0.259	0.955
Block*Age*Workload	0.058	6	0.010	2.103	0.054

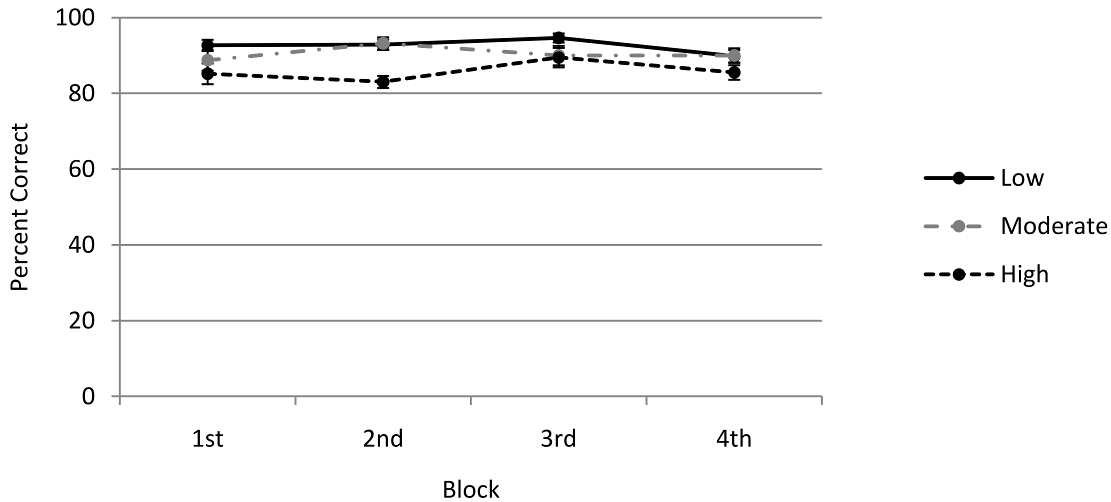


Figure E3. Younger adult’s performance on dispatching trucks across blocks.

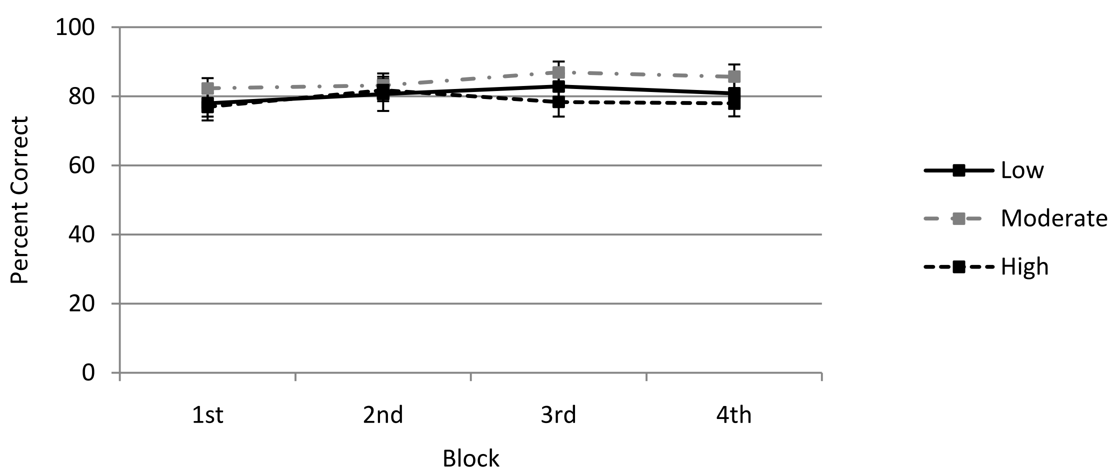


Figure E4. Older adults' performance on dispatching trucks across blocks.

Table E3

Compliance: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	5007.029	1	5007.029	1.792	0.185
Workload	7571.418	2	3785.709	1.356	0.264
Age*Workload	2635.164	2	1317.582	0.471	0.626
Block	1123.821	2.363	475.593	1.732	0.173
Block*Age	883.254	2.363	373.787	1.362	0.259
Block*Workload	1480.526	4.726	313.274	1.141	0.340
Block*Age*Workload	1236.355	4.726	261.608	0.953	0.445

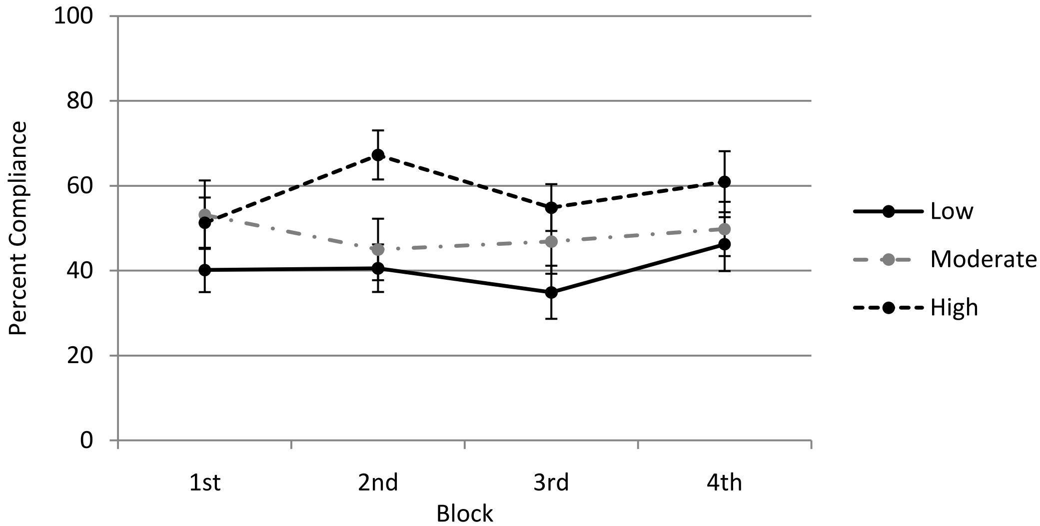


Figure E5. Younger adults' compliance across blocks.

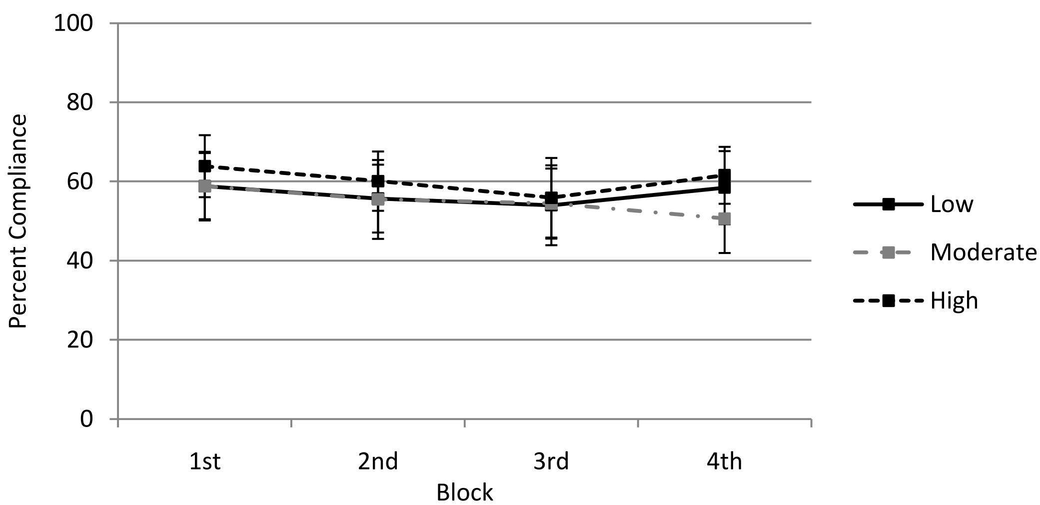


Figure E6. Older adults' compliance across blocks.

Table E4

Reliance: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	470.257	1	470.257	0.312	0.578
Workload	3533.147	2	1766.574	1.172	0.315
Age*Workload	7505.469	2	3752.734	2.489	0.090
Block	1204.967	2.286	527.081	1.909	0.145
Block*Age	364.044	2.286	159.241	0.577	0.585
Block*Workload	860.156	4.572	188.126	0.681	0.625
Block*Age*Workload	1831.882	4.572	400.654	1.451	0.213

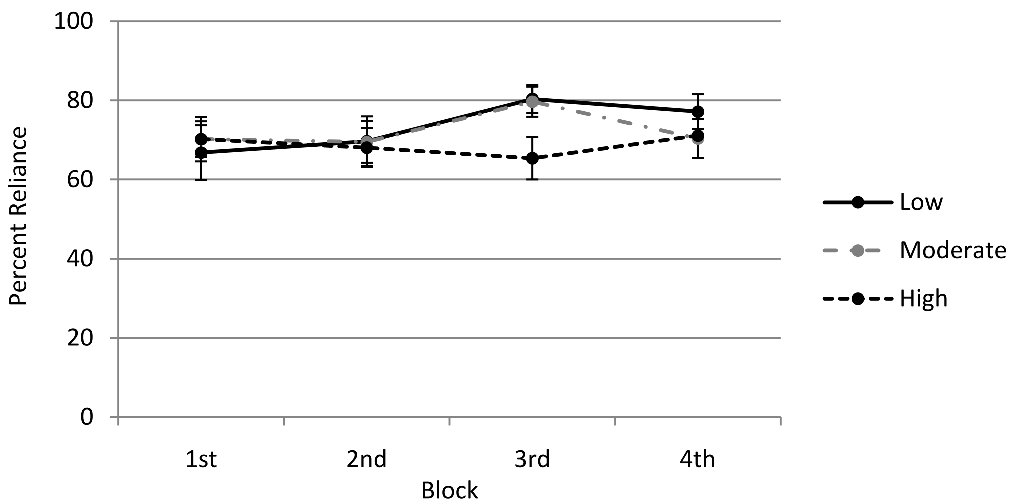


Figure E7. Younger adults’ reliance across blocks.

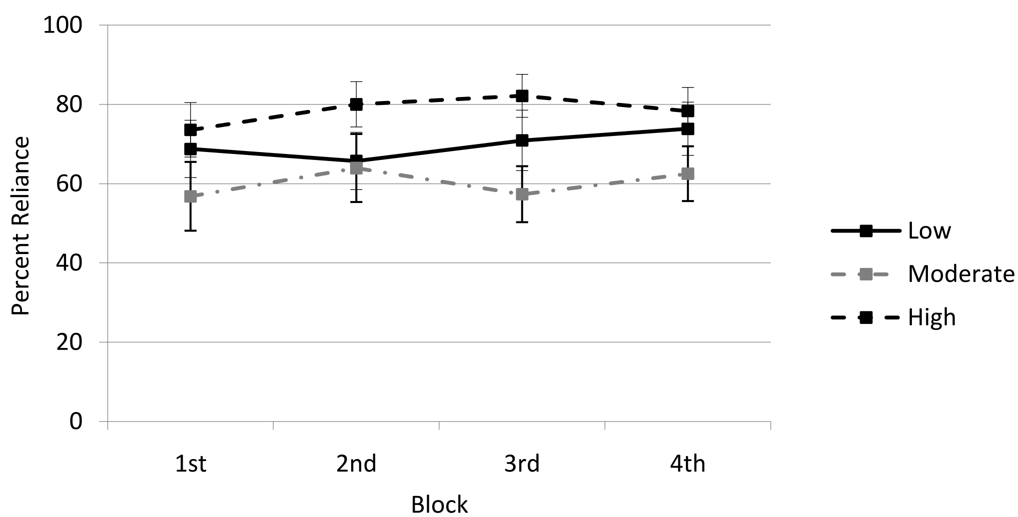


Figure E8. Older adults' reliance across blocks.

Table E5

Trust: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	15.692	1	15.692	5.059	0.027
Workload	8.940	2	4.470	1.441	0.243
Age*Workload	2.040	2	1.020	0.329	0.721
Block	1.051	2.935	0.358	1.094	0.352
Block*Age	1.116	2.935	0.380	1.161	0.325
Block*Workload	5.904	5.871	1.006	3.072	0.007
Block*Age*Workload	2.611	5.871	0.445	1.358	0.234

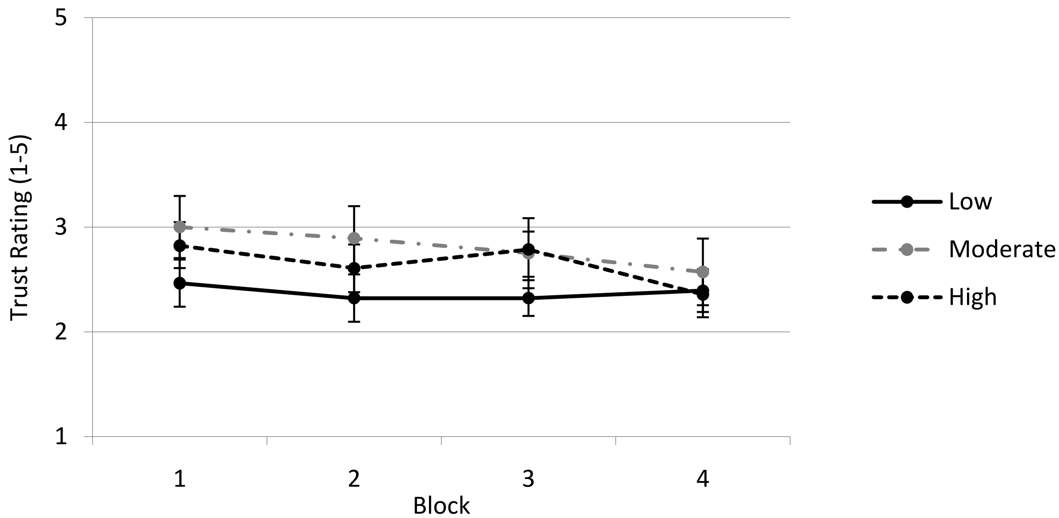


Figure E9. Younger adults' trust rating across blocks.

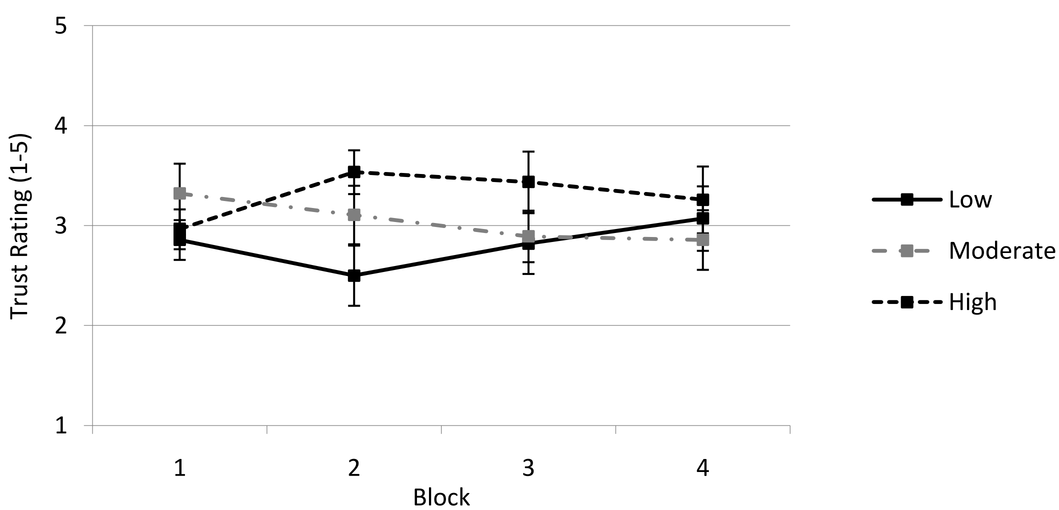


Figure E10. Older adults' trust rating across blocks.

Table E6

Perceived Reliability: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	0.146	1	0.146	1.905	0.171
Workload	0.059	2	0.030	0.388	0.680
Age*Workload	0.268	2	0.134	1.749	0.181
Block	0.006	2.967	0.002	0.168	0.916
Block*Age	0.078	2.967	0.026	2.167	0.093
Block*Workload	0.057	5.933	0.010	0.786	0.580
Block*Age*Workload	0.087	5.933	0.015	1.203	0.306

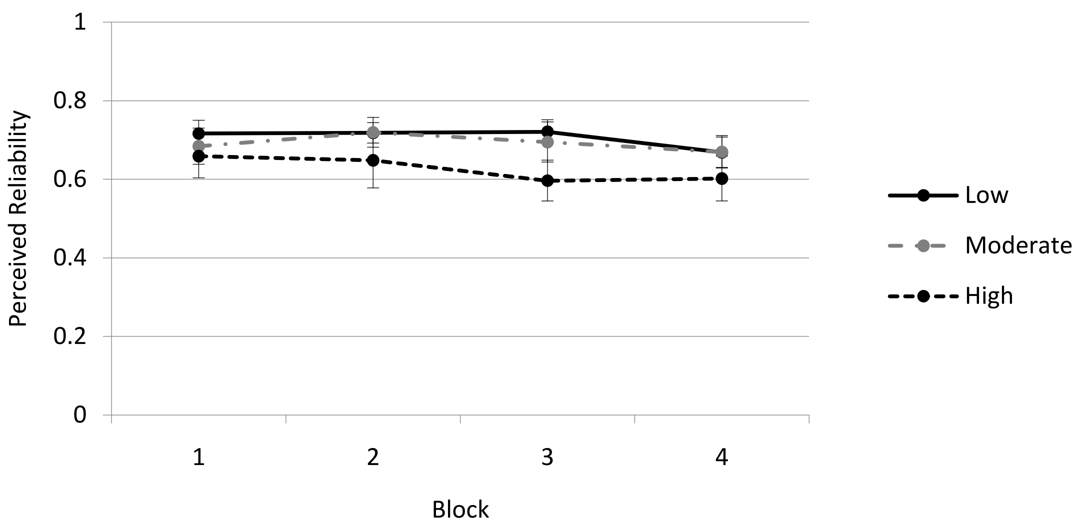


Figure E11. Younger adults' perceived reliability across blocks.

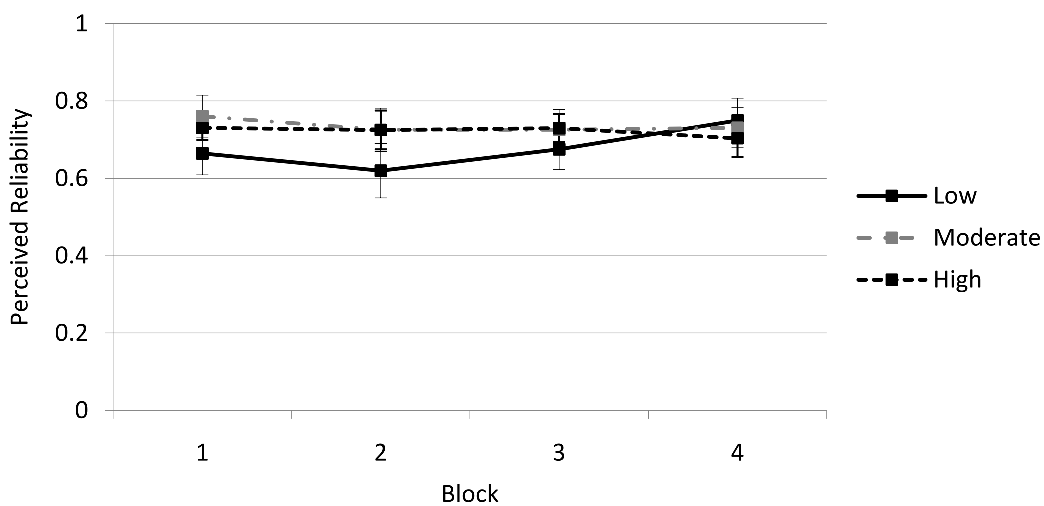


Figure E12. Older adults' perceived reliability across blocks.

Table E7

NASA-TLX: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	12962.345	1	12962.345	9.485	0.003
Workload	9771.277	2	4885.638	3.575	0.033
Age*Workload	5169.606	2	2584.803	1.891	0.158
Block	2286.614	2.418	945.712	10.487	0.000
Block*Age	1252.444	2.418	517.993	5.744	0.002
Block*Workload	297.967	4.836	61.618	0.683	0.632
Block*Age*Workload	696.955	4.836	144.125	1.598	0.165

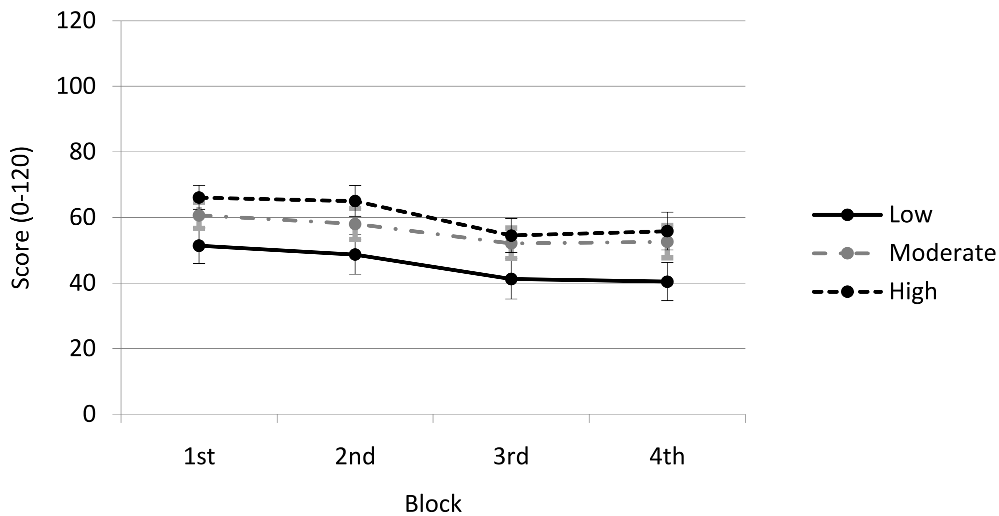


Figure E13. Younger adults' NASA-TLX scores across blocks.

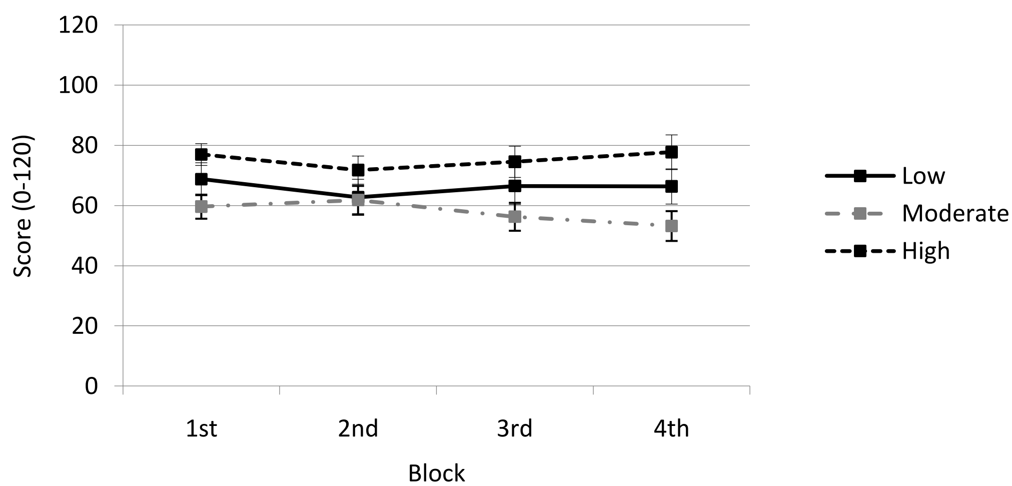


Figure E14. Older adults' NASA-TLX scores across blocks.

APPENDIX F

SWAT DATA

Table F1

SWAT: Repeated Measures ANOVA Summary Table

Source	SS	df	MS	*F*	*p*
Age	112.975	1	112.975	17.119	0.000
Workload	34.667	2	17.334	2.627	0.079
Age*Workload	1.518	2	0.759	0.115	0.892
Block	4.151	2.312	1.795	2.683	0.063
Block*Age	4.632	2.312	2.004	2.995	0.045
Block*Workload	2.621	4.624	0.567	0.847	0.511
Block*Age*Workload	2.750	4.624	0.595	0.889	0.483

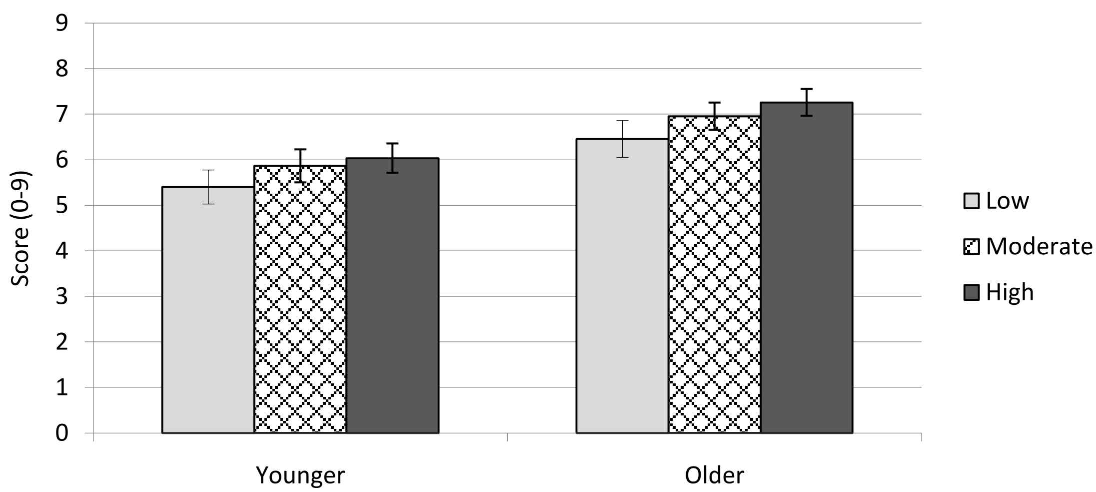


Figure F1. SWAT score by age and workload group.

REFERENCES

Block, R., Zakay, D., & Hancock, P. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology and Aging*, *13*(4), 584-596.

Biros, D.P., Daly, M., & Gunsch, G. (2004). The influence of task workload and automation trust on deception detection. *Group Decision and Negotiation, 13*, 173–189.

Braver, T.S., & West, R. (2008). Working memory, executive control, and aging. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed.) (pp. 189-249). New York, NY US: Psychology Press.

Cepeda, N., Kramer, A., & Gonzalez de Sather, J. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, *37*(5), 715-730.

Dixon, S., Wickens, C., & Chang, D. (2005). Mission Control of Multiple Unmanned Aerial Vehicles: A Workload Analysis. *Human Factors, 47*(3), 479-487.

Dixon, S., & Wickens, C. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors, 48*(3), 474-486.

Dzindolet, M.T., Peterson, S. A., Pomranky, R.A., Pierce, L.G., & Beck, H.P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697-718.

Ezer, N. (2006). Toward an understanding of optimal performance within a human-automation collaborative system: Effects of error and verification costs. Unpublished master's thesis, Georgia Institute of Technology, Atlanta, Georgia.

Fisk, A.D., Rogers, W.A., Charness, N., Czaja, S.J., & Sharit, J. (2004). *Designing for older adults: Principles and creative human factors approaches*. Boca Raton: CRC Press.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (task workload index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). North-Holland: Elsevier Science Publishers B.V.

Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: A review and a new view. *The psychology of learning and motivation: Advances in research and theory, Vol. 22* (pp. 193-225). San Diego, CA US: Academic Press.

Ho, G., Wheatley, D., & Scialfa, C.T. (2005). Age differences in trust and reliance of a medication management system. *Interacting with Computers, 17*(6), 690-710.

Jian, J., Bisantz, A. M., & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4,* 53-71.

Johnson, J.D. (2004). *Type of automation failure: the effects on trust and reliance in automation*. Unpublished master's thesis, Georgia Institute of Technology, Atlanta, Georgia.

Kantowitz, B.H., Hanowski, R.J., & Kantowitz, S.C. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors, 39*(2), 164-176.

Kramer, A., & Madden, D. (2008). Attention. *The handbook of aging and cognition* (3rd ed.) (pp. 189-249). New York, NY US: Psychology Press.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243-1270.

Lee, J.D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*(1), 153-184.

Lee, J.D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors*, *45*(2), 281-295.

Mayer, A.K. (2008). *The manipulation of user expectancies: Effects on reliance, compliance, and trust using an automated system***.** Unpublished master's thesis, Georgia Institute of Technology, Atlanta, Georgia.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44-58.

Muir, B.M. (1994). Trust in automation: I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905-1922.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230-253.

Reid, G., & Nygren, T. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. *Human mental workload* (pp. 185-218). Oxford England: North-Holland.

Sanchez, J., Fisk, A.D., & Rogers, W.A. (2004). Reliability and age-related effects on trust and reliance of a decision support aid. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, CA: Human Factors and Ergonomics Society.

Sanchez, J. (2006). *Factors that affect trust and reliance on an automated aid.* Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta.

Shipley, W.C. (1986). *Shipley Institute of Living Scale*. Los Angeles: Western Psychological Services.

Tsang, P.S., & Vidulich, M.A. (2006). Mental workload and situation awareness. *In G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics: Third Edition (pp. 243-268).*Hoboken, NJ: John Wiley & Sons, Inc.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale III. (3rd Ed.)*. San Antonio, TX: The Psychological Corporation.

Wickens, C. D., & Carswell, C. M. (2006). Information processing. *In G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics: Third Edition (pp. 1570-1596).* Hoboken, NJ: John Wiley & Sons, Inc.

Wilkison, B. (2008). *Effects of mental model quality on collaborative system performance***.** Unpublished master's thesis, Georgia Institute of Technology, Atlanta, Georgia.

Wittmann, M, & Lehnhoff, S. (2005). Age effects in perception of time. *Pscyhological Reports, 97,* 921-935.